

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Análise e Diagramação
Automática de Álbuns Fotográficos

Claudio S. V. da C. Cavalcanti

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Herman Martins Gomes e
José Eustáquio Rangel de Queiroz
(Orientadores)

Campina Grande, Paraíba, Brasil

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Análise e Diagramação
Automática de Álbuns Fotográficos

Claudio S. V. da C. Cavalcanti

Tese submetida à Coordenação do Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Campina Grande - Campus I como parte dos requisitos necessários para obtenção do grau de Doutor em Ciência da Computação.

Herman Martins Gomes e
José Eustáquio Rangel de Queiroz
(Orientadores)

Campina Grande, Paraíba, Brasil



C376a Cavalcanti, Claudio S. V. da C.
Análise e diagramação automática de álbuns fotográficos
/ Claudio S. V. da C. Cavalcanti. - Campina Grande, 2013.
251 f.

Tese (Doutorado em Ciência da Computação) - Universidade
Federal de Campina Grande, Centro de Engenharia Elétrica e
Informática.

1. Diagramação Automática. 2. Algoritmos Genéticos. 3.
Tese - Ciência da Computação. I. Lima, Antonio Gilson
Barbosa de. II. Farias Neto, Severino Rodrigues de. III.
Universidade Federal de Campina Grande - Campina Grande
(PB). IV. Título

CDU 004.932(043)

“ANÁLISE E DIAGRAMAÇÃO AUTOMÁTICA DE ÁLBUNS FOTOGRÁFICOS”

CLAUDIO SEBASTIÃO VASCONCELOS DA CUNHA CAVALCANTI

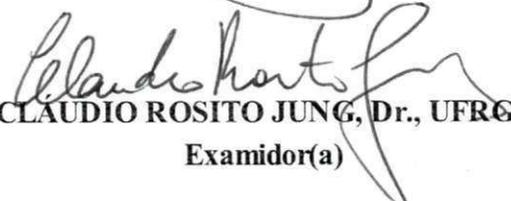
TESE APROVADA EM 27/09/2013



HERMAN MARTINS GOMES, Ph.D, UFCG
Orientador(a)



JOSÉ EUSTÁQUIO RANGEL DE QUEIROZ, D.Sc, UFCG
Orientador(a)



CLAUDIO ROSITO JUNG, Dr., UFRGS
Examidor(a)



LEONARDO VIDAL BATISTA, D.Sc, UFPB
Examinador(a)



JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, D.Sc, UFCG
Examinador(a)



CLÁUDIO DE SOUZA BAPTISTA, Ph.D, UFCG
Examinador(a)

CAMPINA GRANDE - PB

"Dedico este trabalho a Laura, Arthur e John."

Agradecimentos

Um trabalho tão duradouro (quase 5 anos) só poderia ter sido realizado com o apoio de muita gente que, ainda que indiretamente, me ajudaram nesse período a lutar pelo que acreditava. Assim, gostaria de utilizar este espaço para agradecer, por ordem de importância, àqueles que contribuíram com a realização deste trabalho.

Em primeiro lugar, a Deus pela saúde física e mental - e estendo a todas as forças do oculto que desconheço, mas acredito existirem.

Em seguida, aos meus familiares: filho, esposa, mãe, pai, irmãos(ãs), cunhados(as), sobrinhos, tios(as), primos(as) e outros. Sem vocês teria sido impossível.

Agradeço também aos amigos que me ajudaram seja com o apoio (falso ou não, não ligo), seja participando em intermináveis e cansativos experimentos ou me ajudando a codificar, em especial cito (por ordem alfabética, para não ter ciúmes): Carlos, Eanes, Eduardo, Elizabete Fabrício, Larissa, Luciana, Michelly, Xico. Outros também me ajudaram em partes específicas do trabalho e gostaria de deixar o registro: Alysson, Guilherme, Rafael e Talita.

Agradeço aos antigos mas eternos amigos recifenses que ainda que pelo facebook me ajudaram de alguma forma.

Também agradeço aos orientadores pela orientação e pelo apoio na realização desta pesquisa. Estendo aqui também a todos os demais professores e funcionários da COPIN. Agradeço aos colegas do Analytics, tanto os professores quanto os participantes, por terem me ajudado e terem acreditado em mim.

Agradeço aos colegas da UAAMI, Joelson, Max, Róbson, Mayra, Francieudo e outros - são muitos - por terem me deixado escapar do mundo real e realizar alguns sonhos.

Agradeço aos membros da banca examinadora pelos comentários.

Agradeço aos ex-alunos e professores da UEPB, IFPB e UFCG que me apoiaram e me desafiaram.

Por fim, se você não se enquadrou em nenhum dos itens acima, mas está lendo essa seção é porque, certamente, foi muito importante para mim, obrigado a você também!!!

Resumo

A contínua redução dos custos de dispositivos de aquisição de imagens e a consequente popularização da fotografia digital tem suscitado uma demanda por ferramentas para auxiliar no processo de seleção e organização de fotografias em grandes coleções. Esta tese insere-se nesse contexto, tendo como foco a concepção de métodos automáticos para a análise, a seleção, o agrupamento e a posterior diagramação de álbuns fotográficos. Com o objetivo de selecionar fotografias, são propostos e discutidos métodos para a análise da composição fotográfica. Em particular, é proposto um método para a análise do espaço negativo. Investiga-se também a análise da harmonia das cores de uma fotografia. Para tanto, é proposto um método para a segmentação de cores de imagens, obtido por intermédio de uma estratégia de computação evolucionária. Para a realização do agrupamento, propõe-se a utilização de meta-informações dos horários em que as fotografias foram obtidas e da distribuição de cores de cada imagem. No tocante ao processo de diagramação de álbuns fotográficos, é proposto um método para o recorte automático de fotografias digitais, com base em seu conteúdo. Tal método utiliza-se de uma série de características (e.g., presença da pele humana e faces detectadas) para a definição de regiões da fotografia com conteúdo de maior saliência ou relevância, as quais deverão ser preservadas. Foi realizada uma avaliação experimental subjetiva com o intuito de avaliar os métodos propostos. Os resultados permitiram verificar que os métodos propostos para a análise de fotografias produziram resultados consistentes com a opinião humana. Também se pôde constatar que a estratégia de computação evolucionária utilizada, mesmo fazendo uso de uma pequena quantidade de métodos destinados ao processamento de imagens, produziu um segmentador de imagens consistente com o objetivo almejado. Enfim, após integradas as partes e gerado o sistema completo de diagramação, foi realizado um experimento final, no qual 81% dos participantes concordaram que tal sistema auxilia significativamente na produção de um álbum fotográfico.

Abstract

The popularization of digital photography, due to the continued decline in the costs of photographic equipments, has resulted in a demand for tools and methods to facilitate the selection and organization of very large collections of photographs. In such a context, this thesis aims at conceiving methods for analysis, selection, clustering, and layouting of photo albums. Aiming the photo selection, methods for photographic composition analysis are proposed. In particular, it is proposed a method for analyzing the negative space of a photograph. It is also investigated the analysis of the color harmony of a photograph. Such analysis required an image segmentation algorithm, which was developed by adopting an evolutionary computation strategy. Photo clustering is achieved by the use of photo timestamps obtained from images metadata. Additionally, the image color distribution is used in order to provide a content-based photo clustering. For the layouting process of digital photo albums, it is proposed a method for content-aware cropping of digital photographs. Such method uses a set of features (e.g., human skin-tone and detected face coordinates) for defining areas with salient regions, which should be preserved. A subjective evaluation was performed in order to evaluate the proposed methods. The experimental results show that the proposed method is promising because they are correlated with human decisions. It was also noted that the employed evolutionary method produced an image segmentation algorithm which is adequate for its purposes, even using a small set of image processing tools. Finally, a final experiment was performed, with the whole system integrated, and it was shown that 81% of the participants agreed that it significantly reduces the effort in producing a photo-book.

Lista de símbolos e abreviaturas

AZ&C	<i>Automatic zoom & crop method</i>
DFD	Diagrama de fluxo de dados
ED	Detector de bordas (<i>edge detector</i>)
FD	Detector de faces (<i>face detector</i>)
FPS	Quadros por segundo (<i>frames per second</i>)
GA ou AG	Algoritmos genéticos (<i>genetic algorithms</i>)
GP ou PG	Programação genética (<i>genetic programming</i>)
GNA	Gerador de números aleatórios
HSV	Espaço de cor matiz-saturação-intensidade (<i>hue-saturation-value</i>)
HTML	Linguagem de marcação de hipertextos (<i>Hypertext markup language</i>)
IA ou AI	Inteligência artificial (<i>artificial intelligence</i>)
IR	Recuperação de informação (<i>information retrieval</i>)
LAB	Espaço de cor CIE L*a*b
MP	<i>Mega pixel</i> , o equivalente a 1 milhão de <i>pixels</i>
MSCM	<i>Multiple saliency cropping method</i>
PCA	Análise de componentes principais (<i>principal component analysis</i>)
PDF	Formato de documento portátil (<i>portable document format</i>)
RDI	Região de interesse
RGB	Espaço de cor vermelho-verde-azul (<i>Red-Green-Blue</i>)
RIBC	Recuperação de imagens com base no conteúdo
RNA	Redes neurais artificiais
SRQ	Soma residual dos quadrados
SST	Soma dos efeitos

- STD Detector de pele (*skin-tone detector*)
- SVN Máquinas de vetor de suporte (*support vector machines*)
- UML Linguagem de modelagem unificada (*unified modeling language*)
- VA Atenção visual (*visual attention*)
- VDP Impressão de dados variáveis (*variable data printing*)
- XML Linguagem de marcação extensível (*extensible markup language*)

Lista de Figuras

1.1	Sistema proposto.	2
2.1	Regra dos terços: um dos olhos está sobre um dos quatro pontos dos terços, enquanto o outro está sobre a linha dos terços horizontal superior.	17
2.2	Ilustração dos níveis de <i>zoom</i> : (a) longa distância; (b) média distância; (c) <i>close-up</i> ; (d) <i>close-up</i> extremo.	20
2.3	Exemplo de violação da integridade: topo e laterais da cabeça foram cortados.	24
2.4	Padrões de distribuição dos níveis de contraste. Fonte: adaptado de Cohen-Or et al. (2006).	27
2.5	Padrões de harmonia de cores usados por diversos profissionais de <i>design</i>	28
2.6	Ilustração dos princípios básicos de um projeto gráfico.	42
2.7	Exemplo de distorção causada por métodos de redirecionamento.	55
3.1	Diagrama geral do sistema proposto.	57
3.2	Ilustração do sistema proposto.	59
4.1	Histogramas de projeção médios das fotografias que, respectivamente, obedecem à regra do Espaço Negativo e não obedecem à regra do Espaço Negativo.	68
4.2	Histogramas de projeção médios e simétricos das fotografias que, respectivamente, obedecem à regra do Espaço Negativo e não obedecem à regra do Espaço Negativo.	68

4.3	Densidade de probabilidade das fotografias consideradas com alta qualidade (em azul) e das consideradas de baixa qualidade (em vermelho) em cada uma das métricas utilizadas.	75
5.1	Modelo proposto nesta tese de organização hierárquica pelo tempo. . .	81
5.2	Histograma tridimensional de uma imagem.	84
6.1	Diagrama de atividades representando o método recortar da classe <i>MSCM</i>	96
6.2	Saídas para cada extrator de características: (a) a imagem original, (b) o FD com regiões antropométricas, (c) a VA (d) as bordas do ED, (e) STD, e (f) a soma das características.	100
6.3	Síntese gráfica do algoritmo de recorte.	106
7.1	Tela de captura da página <i>Web</i> usada para obter a opinião dos participantes sobre a qualidade dos recortes realizados, ordenando-os do melhor (1) para o pior (4).	114
7.2	Exemplo da saída dos métodos <i>MSCM</i> e do <i>AZ&C</i> : em (a) o <i>MSCM</i> foi considerado melhor do que o <i>AZ&C</i> , em (b) o <i>MSCM</i> foi considerado pior pelos participantes.	118
7.3	Página <i>Web</i> empregada para capturar a opinião dos participantes sobre a qualidade dos recortes realizados.	121
7.4	Gráfico em barras ilustrando a preferência dos métodos (em percentuais)	122
7.5	Exemplos de imagens da base de treinamento, as respectivas imagens de referência e as imagens geradas pelo sistema resultante do método proposto.	128
7.6	Exemplos de imagens da base de testes, geradas a partir da espécie vencedora.	128
7.7	Apesar da região vermelha ter área pequena, comparado com o restante da imagem, possui grande impacto na percepção.	130
7.8	Cubo de cores indicando a cor considerada mais importante por todos os participantes do experimento.	132
7.9	Curvas ROC	135

7.10	Intervalos de confiança da medida-F dos resultados de seleção de um usuário para os demais.	141
7.11	Diagrama de caixa com os valores da medida-f dos quatro métodos propostos.	142
7.12	Diagrama de caixa com os valores da medida-f dos quatro métodos propostos, considerando-se apenas os usuários com a média da medida F superior a 0,10.	143
7.13	Diagrama de caixa com os valores da medida-f dos quatro métodos propostos, considerando-se apenas os usuários com a média da medida F superior a 0,20.	143
7.14	Diagrama de caixa com os valores da medida-f dos quatro métodos propostos, considerando-se apenas os usuários com a média da medida F superior a 0,30.	144
B.1	Diagrama do método de computação evolutiva utilizado.	193
B.2	Modelo de processamento sequencial utilizado.	195
D.1	Captura da tela de participação do experimento sobre cores.	207
D.2	Distribuição dos erros de rotulagem dos usuários	210
D.3	Distribuição do número de rotulagens por imagem.	211
D.4	Histograma do número de imagens pelo número de participantes que rotularam tal quantidade.	211
D.5	Distribuição da idade dos participantes do experimento	212
E.1	Imagens utilizadas no experimento (1/7).	216
E.2	Imagens utilizadas no experimento (2/7).	217
E.3	Imagens utilizadas no experimento (3/7).	218
E.4	Imagens utilizadas no experimento (4/7).	219
E.5	Imagens utilizadas no experimento (5/7).	220
E.6	Imagens utilizadas no experimento (6/7).	221
E.7	Imagens utilizadas no experimento (7/7).	222

F.1	Exemplos de imagens da base de treinamento, imagens de referência correspondentes e imagens geradas pelo sistema treinado.	223
F.2	Exemplos de imagens da base de treinamento, imagens de referência correspondentes e imagens geradas pelo sistema treinado.	224
F.3	Exemplos de imagens da base de treinamento, imagens de referência correspondentes e imagens geradas pelo sistema treinado.	224
F.4	Exemplos de imagens produzidas e cores extraídas.	226
F.5	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	227
F.6	Exemplos de imagens produzidas e cores extraídas.	228
F.7	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	229
F.8	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	230
F.9	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	231
F.10	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	232
F.11	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	233
F.12	Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.	234
G.1	Imagens utilizadas no experimento (1/21).	236
G.2	Imagens utilizadas no experimento (2/21).	237
G.3	Imagens utilizadas no experimento (3/21).	238
G.4	Imagens utilizadas no experimento (4/21).	239
G.5	Imagens utilizadas no experimento (5/21).	240
G.6	Imagens utilizadas no experimento (6/21).	240
G.7	Imagens utilizadas no experimento (7/21).	241
G.8	Imagens utilizadas no experimento (8/21).	242

G.9	Imagens utilizadas no experimento (9/21).	242
G.10	Imagens utilizadas no experimento (10/21).	243
G.11	Imagens utilizadas no experimento (11/21).	244
G.12	Imagens utilizadas no experimento (12/21).	245
G.13	Imagens utilizadas no experimento (13/21).	246
G.14	Imagens utilizadas no experimento (14/21).	246
G.15	Imagens utilizadas no experimento (15/21).	247
G.16	Imagens utilizadas no experimento (16/21).	247
G.17	Imagens utilizadas no experimento (17/21).	248
G.18	Imagens utilizadas no experimento (18/21).	249
G.19	Imagens utilizadas no experimento (19/21).	249
G.20	Imagens utilizadas no experimento (20/21).	250
G.21	Imagens utilizadas no experimento (21/21).	250

Lista de Tabelas

2.1	Parâmetros para a regra do <i>zoom</i>	20
2.2	Proporções de cor definidas por Itten (1961).	30
4.1	Fontes de busca e número de imagens adquiridas por semente.	67
6.1	Pesos obtidos pelos extratores de características utilizados.	105
7.1	Resumo do desempenho do método.	115
7.2	Resultados do teste de Tukey-Kramer.	115
7.3	Distribuição dos tipos de fotografia usados no experimento estendido.	120
7.4	Método da Somatória de Valores (LIKERT, 1932) aplicado aos resultados da pesquisa envolvendo 100 imagens e 12 participantes.	122
7.5	Método da Somatória de Valores (LIKERT, 1932) aplicado aos resultados da pesquisa envolvendo 94 imagens e 12 participantes.	122
7.6	Comparação direta entre os métodos. Cada linha mostra o número de vezes que o recorte realizado por um método foi considerado superior ao recorte realizado pelo método indicado na coluna.	123
7.7	Valor absoluto e percentual de votos nos quais a imagem gerada pelo MSCM foi considerada igual ou superior às imagens geradas pelos métodos empregando características individualmente.	123
7.8	Resultados do Teste de Comparação Pareada de Tukey-Kramer entre o MSCM e os métodos que utilizam características individuais.	125
7.9	Agrupamento realizado pela aplicação do Teste de Tukey-Kramer.	125
A.1	Associação entre fatores e níveis aos rótulos que serão utilizados.	172
A.2	Resultados dos experimentos dadas as combinações.	173

A.3	Tabela de Sinais para o problema. A, B e C indicam os fatores e as possíveis combinações. Y o percentual de acerto para cada uma das combinações e I o sinal para a soma de todos os efeitos.	173
B.1	Funções e respectivas quantidades de parâmetros.	186
C.1	Tabela comparativa de preços de câmeras digitais da marca Nikon, entre 1999 e 2010, conforme registrados por Rockewll (2011).	200
C.2	Tabela comparativa dos preços de câmeras digitais da família D100 a D300s, conforme registrados por Rockewll (2011).	201
D.1	Perfil fotográfico do participante.	213
D.2	Identificador, média e desvio padrão das rotulagens das piores fotografias. As notas variam entre 0 e 5.	214
D.3	Identificador, média e desvio padrão das rotulagens das melhores fotografias. As notas variam entre 0 e 5.	215

Conteúdo

1	Introdução	1
1.1	Motivações	4
1.2	Objetivos	5
1.2.1	Objetivo geral	6
1.2.2	Objetivos específicos	6
1.3	Metodologia adotada	7
2	Fundamentação teórica	8
2.1	Algoritmos para a análise de fotografias	8
2.1.1	Antropometria	15
2.1.2	Regra dos terços	16
2.1.3	Regra do <i>zoom</i>	18
2.1.4	Regra da integridade	22
2.1.5	Análise da harmonia das cores de uma imagem	24
2.1.6	Extração de informações	31
2.1.7	Agrupamento	34
2.1.8	Discussão	35
2.2	Diagramação automática de documentos	40
2.2.1	Fundamentação sobre diagramação	40
2.2.2	Pesquisas sobre diagramação automática de documentos	44
2.2.3	Discussão	50
2.3	Recorte automático de fotografias	51
2.3.1	Algoritmos de recorte	52
2.3.2	Algoritmos para redirecionamento de imagens	53

2.3.3	Discussão	55
2.4	Considerações Finais	56
3	Sistema proposto	57
3.1	Visão geral do projeto	57
3.2	Formalização do problema	60
3.3	Considerações finais	62
4	Métodos para a análise de fotografias digitais	64
4.1	Regras de composição fotográfica: proposta da regra do espaço negativo	65
4.2	Análise da harmonia das cores	70
4.2.1	Segmentação	70
4.2.2	Cálculo da harmonia	73
4.3	Combinação dos avaliadores	73
4.4	Considerações finais	76
5	Agrupamento de fotografias	78
5.1	Função de similaridade proposta	79
5.1.1	Similaridade cronológica	80
5.1.2	Similaridade por conteúdo	83
5.2	Distribuição das fotografias ao longo das páginas do álbum fotográfico .	85
5.3	Considerações finais	88
6	Método para a diagramação automática de álbuns fotográficos	89
6.1	Extensão da abordagem existente	91
6.2	Recorte automático de fotografias	93
6.2.1	Introdução	93
6.2.2	Método de recorte usando múltiplas saliências (<i>multiple saliency cropping method - MSCM</i>)	95
6.2.3	Extratores de características	97
6.3	Integração do sistema de recorte ao diagramador automático	109
6.4	Considerações finais	111

B.1.5	Algoritmo de segmentação	189
B.1.6	Estratégia para agrupamento a partir do espaço de cor HSV . . .	190
B.2	Aspectos computacionais	190
B.3	Adaptação da biblioteca GALib	191
B.4	Considerações finais	196
C	Evolução da tecnologia de captura de fotografias	197
C.1	Evolução da tecnologia das câmeras digitais	198
C.2	Evolução dos preços das câmeras digitais	199
C.3	Mudanças no hábito de fotografar e no processo pós-fotografia	202
C.4	Conclusão	203
D	Questionário sobre os hábitos fotográficos	204
D.1	Transcrição dos questionários realizados no experimento	204
D.2	Sumário da análise realizada sobre os dados resultantes do experimento	209
D.2.1	Análise do perfil dos participantes	212
D.2.2	Análise das notas atribuídas às imagens	214
D.3	Considerações finais	215
E	Fotografias utilizadas no experimento de rotulagem de cores	216
F	Fotografias utilizadas no treinamento do segmentador	223
G	Diário do evento	235
G.1	Diário do evento	236
G.2	Considerações finais	251

Capítulo 1

Introdução

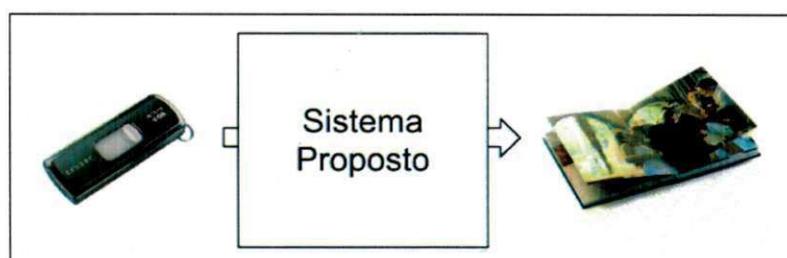
As fotografias digitais são cada vez mais comuns em aplicações as mais diversas, graças à grande quantidade de dispositivos de captura de fotografias existentes. Além das tradicionais e cada vez mais baratas câmeras fotográficas digitais, pode-se também capturar fotografias a partir de telefones celulares, dispositivos móveis como os *tablets* e câmeras não portáteis tais como as *webcam*¹. Uma consequência da facilidade de acesso a tal diversidade de dispositivos tem sido a produção de uma quantidade crescente de fotografias. Apesar de não ter sido identificado, no âmbito desta pesquisa, nenhum estudo publicado sobre o destino que os consumidores dão a suas fotografias digitais, percebeu-se por meio de estudos informais realizados, que grande parte delas não chega a ser publicada, seja por meio físico, após a impressão, seja por meio digital, em *sites* especializados, redes sociais ou outro meio (e.g., álbum em *desktop* ou *tablet*).

Objetivou-se, no âmbito desta pesquisa de doutorado, desenvolver um sistema destinado a dar suporte aos seguintes processos: (1) a seleção de fotografias de interesse e (2) a estruturação diagramada das fotografias de interesse em álbuns fotográficos. A presente proposta de pesquisa constitui uma extensão da investigação realizada por Cavalcanti (2007), sobre o problema da composição automática de fotografias, na qual se condicionavam fotografias ao maior número possível de regras de composição fotográfica sem que sofressem alterações significativas.

¹No Apêndice C, é apresentado um breve estudo sobre a evolução da tecnologia de captura de fotografias e a queda dos preços dos dispositivos de captura nos últimos 10 anos.

Na Figura 1.1, ilustra-se o comportamento geral planejado para o sistema proposto. Dado um conjunto de fotografias digitais, a partir de uma mídia qualquer (aqui representada por um dispositivo portátil de memória *flash*), o sistema seleciona automaticamente as fotografias que comporão um documento diagramado automaticamente, e.g., um livro de recortes, um sítio *Web*, uma revista. A modelagem detalhada deste sistema é descrita no Capítulo 3.

Figura 1.1: Sistema proposto.



Na abordagem adotada por Cavalcanti (2007), fotografias são alteradas de modo a obedecerem regras de composição fotográficas clássicas. Entretanto, tais alterações não necessariamente as qualificam para fins de impressão (e/ou apresentação em tela), apesar de terem sido mostrados, no mesmo trabalho, indícios experimentais de que a utilização de tais regras aumenta a aceitação das fotografias alteradas.

Apesar de existirem regras e técnicas que podem tornar mais objetiva a análise de fotografias, sua avaliação frequentemente possui um cunho elevado de subjetividade, uma vez que o julgamento do observador pode ser influenciado pelos seguintes aspectos elencados nesta pesquisa:

- **Os alvos humanos na fotografia.** O observador tende a preferir fotografias nas quais aparecem alvos humanos que fazem parte de seu núcleo de amigos;
- **O local onde a fotografia foi obtida.** O observador tende a preferir fotografias de locais que lhe trazem boas lembranças;
- **Fisionomia.** O observador tende a preferir fotografias contendo fisionomias que mais lhe agradem; e

- **Outras preferências individuais.** O observador pode possuir preferências relativas aos aspectos temporais (eg. fotografias antigas vs. recentes, diurnas vs. noturnas etc) e aos aspectos visuais (cores ou texturas preferenciais) dentre outras.

Apesar de alguns itens, dentre aqueles listados anteriormente, poderem ser considerados objetivos (i.e., o local onde a fotografia foi obtida e aspectos temporais), estes itens são, na verdade, considerados subjetivos do ponto de vista individual, pois sua influência na avaliação de uma fotografia é fortemente dependente de aspectos individuais (e.g., enquanto uma fotografia feita diante da torre Eiffel pode ter valor sentimental para um turista brasileiro, para um parisiense pode parecer uma paisagem comum e corriqueira). A inclusão de tais aspectos subjetivos implica a necessidade de incorporação de técnicas de análise capazes de interpretar os diversos objetos presentes na cena e seus relacionamentos. Este é um tópico de investigação ainda não conclusivo na literatura.

Uma vez que, de acordo com a pesquisa bibliográfica realizada, não foram encontrados métodos e técnicas capazes de incorporar tal nível de subjetividade, no processo de avaliação adotado nesta pesquisa foram incorporadas restrições quanto ao tipo de fotografia considerado e aos casos nos quais as seguintes considerações são válidas:

- As fotografias encontram-se em formato digital;
- As fotografias foram obtidas por fotógrafos amadores;
- Não há distinção entre os alvos humanos presentes nas fotografias;

O presente documento está estruturado em seis capítulos. O restante deste capítulo contempla as informações descritas a seguir. Na Seção 1.2, são descritos os objetivos geral e específicos desta pesquisa. Em seguida, na Seção 1.3, é descrita a metodologia usada na pesquisa relatada neste documento. Os demais capítulos estão organizados como segue. O Capítulo 2 contém uma revisão da literatura técnica da área, na qual se descrevem e se discutem abordagens e técnicas pertinentes ao contexto da pesquisa proposta. No Capítulo 3, é descrito o sistema proposto nesta pesquisa. No

Capítulo 4, é detalhado o sistema destinado à avaliação de fotografias, o qual é composto por módulos de avaliação da composição fotográfica, em particular pela análise do espaço negativo de uma fotografia e de sua harmonia cromática. Em seguida, no Capítulo 5, é descrito um método proposto nesta pesquisa para o agrupamento das fotografias previamente selecionadas. No Capítulo 6, é apresentado o método proposto nesta pesquisa para a diagramação de documentos, para o qual foi desenvolvido um módulo de recorte automático de fotografias, de forma a ajustá-las a proporções mais adequadas à diagramação, permitindo maior utilização da área disponível na página, o qual também é descrito no mesmo capítulo. A seguir, no Capítulo 7, descrevem-se os experimentos realizados, com o intuito de validar as abordagens propostas ao longo desta tese. Finalmente, no Capítulo 8, são apresentadas as considerações finais sobre a pesquisa ora descrita e sugeridos temas para investigações futuras.

1.1 Motivações

A obtenção de fotografias digitais tem tido diversos adeptos, dentre outros motivos, em virtude do barateamento dos equipamentos e o aumento na capacidade de armazenamento. De acordo com Huynh et al. (2005), em uma pesquisa realizada em 2003, 17% dos usuários de câmeras fotográficas obtêm, em um período aproximado de cinco anos, de 3.000 a 6.000 fotografias. No Apêndice C, há uma discussão mais aprofundada sobre a evolução dos equipamentos de captura e possíveis tendências do campo da fotografia.

Tendo em mãos quantidades de fotografias dessa ordem, é de se esperar que grande parte dos usuários não tenha planos específicos para a sua utilização, seja por não disporem de tempo ou por não possuírem habilidade para a sua manipulação.

Abordagens populares têm se fundamentado na criação de álbuns digitais de fotografias, os fotolivros. Os fotolivros constituem uma evolução dos antigos álbuns digitais, nos quais os proprietários dispunham as fotografias de forma geralmente sequencial, posicionando uma fotografia por página do álbum. Entretanto, a geração de fotolivros apresenta um nível de dificuldade alto, uma vez que requer o conhecimento prévio de alguns princípios de diagramação. Uma pesquisa realizada em 2008 pela PMA² mostrou

²PMA é uma associação internacional de profissionais da área de imagem, acessível pelo endereço Web <http://www.pmai.org/>

que 47,7% dos indivíduos que iniciaram a construção de um fotolivro não o concluíram, em virtude do tempo demandado e do nível de dificuldade do processo (EGOROVA; SAFONOV, 2011; RESEARCH, 2008).

A dificuldade de construção de fotolivros poderia ser eliminada ou, ao menos, reduzida caso fosse utilizada uma abordagem fundamentada em métodos automáticos para a seleção, o agrupamento e a diagramação das fotografias de interesse. Tal abordagem pode impactar nas indústrias de processamento e impressão fotográfica, uma vez que reduz drasticamente o esforço necessário para a produção de fotolivros ou álbuns fotográficos.

Por fim, destaca-se como motivação adicional o fato da área objeto de estudo desta tese ser pouco explorada na literatura especializada e apresentar um elevado grau de dificuldade para a produção de soluções, uma vez que exige compreensão de alto nível do conteúdo da fotografia e do histórico dos eventos ao qual estas fotografias estão inseridas, não havendo pesquisas na área da visão computacional que apresentem resultados conclusivos para tal nível de conhecimento.

1.2 **Objetivos**

A partir da breve descrição do problema, realizada no início deste capítulo, pode-se formular a seguinte questão de pesquisa: É possível gerar automaticamente álbuns fotográficos, a partir de um conjunto não rotulado de fotografias, de modo que os resultados sejam melhores do que aqueles produzidos pelos métodos atualmente existentes?

Partindo-se da questão previamente apresentada, pretende-se confirmar ou invalidar as seguintes hipóteses:

1. A utilização de algoritmos de análise da composição fotográfica podem permitir a realização da seleção automática de fotografias com resultados compatíveis àqueles produzidos por humanos;
2. A utilização da cronologia no processo de organização de um álbum fotográfico faz com que os usuários o prefiram;
3. A partir da utilização de algoritmos genéticos, é possível produzir uma variação do

algoritmo BRIC (ATKINS, 2004) cujos resultados sejam melhores do que aqueles obtidos a partir de tal algoritmo;

A partir destas hipóteses, foram definidos os objetivos geral e específicos desta tese, conforme explicitados nas Seções 1.2.1 e 1.2.2:

1.2.1 **Objetivo geral**

O objetivo desta pesquisa é desenvolver um sistema para a diagramação automática de álbuns fotográficos, a partir de um conjunto não classificado de fotografias, as quais serão selecionadas, agrupadas e, em seguida, posicionadas no álbum, a partir da utilização de métodos destinados à avaliação do nível de proximidade existente entre estas.

1.2.2 **Objetivos específicos**

Para que o objetivo geral seja atingido, fazem-se necessários os seguintes objetivos específicos:

- desenvolver algoritmos para a análise da qualidade fotográfica, à luz de um conjunto de regras de composição fotográfica, em particular, regras para a análise do espaço negativo e da harmonização das cores dos elementos presentes em regiões de uma fotografia digital;
- desenvolver um algoritmo para o cálculo da similaridade entre fotografias de modo a servir como parâmetro para a realização do agrupamento de tais fotografias;
- definir uma função destinada a priorizar a exibição das fotografias preferidas pelo usuário;
- desenvolver um algoritmo que posicione um conjunto de fotografias em uma página, de modo a maximizar a área utilizada, dado que nesta pesquisa optou-se por um estilo de leiaute que visa à utilização plena da área disponível.

Para atingir-se tais objetivos, duas poderiam ser as abordagens: automática ou semi-automática. Optou-se, nesta pesquisa, pela abordagem automática, uma vez que (1) este é um passo para a abordagem semi-automática, dado que para se apresentar um

resultado parcial a ser modificado pelos usuários do sistema, é essencial reduzir o esforço para os ajustes que estes realizarão; e (2) a abordagem automática gera um álbum fotográfico mais rapidamente e com menos esforço que a abordagem semi-automática, dado que já foi apontado, na Seção 1.1, que diversos usuários não concluíram a produção do álbum fotográfico, em virtude do tempo demandado.

1.3 Metodologia adotada

Nesta seção, é descrita a metodologia que fundamentou a condução da presente pesquisa.

Inicialmente, foi realizada uma revisão bibliográfica dos temas inseridos no âmbito desta tese, a saber: harmonização de cores, composição fotográfica, autodiagramação, aprendizagem dinâmica e métodos para a extração de características. Tal revisão bibliográfica teve como objetivo realizar um levantamento do estado da arte das áreas supramencionadas, bem como identificar possíveis lacunas de pesquisa.

Após a revisão bibliográfica, foi definida a arquitetura do sistema e especificadas as possíveis inovações desejáveis em cada técnica estudada.

Em seguida, os módulos do sistema foram desenvolvidos. Tais módulos foram desenvolvidos separadamente e, em seguida, integrados em um sistema final. A implementação de cada módulo foi acompanhada de testes e experimentos intermediários, com fins a sua validação. Na maior parte dos módulos, foram utilizadas métricas subjetivas para a avaliação, uma vez que grande parte do sistema desenvolvido não é passível de avaliação objetiva. Em seguida, realizou-se um experimento com todo o sistema integrado, no qual foram avaliadas as opiniões dos utilizadores do sistema do qual derivaram-se os resultados.

Por fim, foram publicados os resultados produzidos pelos módulos de recorte automático e de análise de fotografias e, por fim, os resultados da pesquisa bibliográfica realizada. Tais publicações são descritas em mais detalhes na Seção 8.2.

Capítulo 2

Fundamentação teórica

Neste capítulo, é apresentada uma revisão bibliográfica associada aos principais problemas abordados na pesquisa. Adicionalmente, durante a revisão, buscou-se evidenciar os principais fundamentos e conceitos norteadores da pesquisa.

A compreensão prévia de tais conceitos e abordagens descritas na bibliografia revisada possibilitará melhor entendimento do restante deste documento. São tratadas as principais etapas da pesquisa, partindo da análise de fotografias até a diagramação do documento. Na Seção 2.1, é apresentado um estudo sobre os algoritmos destinados à avaliação de fotografias. Na Seção 2.2, é realizado um estudo sobre os métodos destinados à realização da diagramação de documentos, em particular, álbuns fotográficos. Na Seção 2.3, são apresentadas as pesquisas destinadas ao recorte ou redirecionamento de fotografias. Por fim, na Seção 2.4, apresentam-se as considerações finais deste capítulo.

2.1 Algoritmos para a análise de fotografias

Nesta seção, são revisadas pesquisas relevantes sobre análise de fotografias. A partir do levantamento bibliográfico realizado nesta pesquisa, foram identificadas três linhas de investigação, as quais foram utilizados nesta seção para refletir tal levantamento:

1. **Ordenação por importância.** O objetivo é aplicar um escore para cada fotografia segundo uma escala pré-definida (e.g., de zero a dez, boa ou ruim) de acordo com algum critério de interesse (e.g., a fotogenia de alvos humanos, a qualidade do recorte);

2. **Extração de informações.** O objetivo é detectar a presença e a posição de elementos de interesse (e.g., alvos humanos e faces); e
3. **Agrupamento.** O objetivo é agrupar fotografias, por intermédio de alguma característica extraída, (e.g., dia e hora, local, faces reconhecidas).

Algoritmos destinados à ordenação por importância utilizam métricas para atribuir escores às fotografias, permitindo a ordenação das fotografias com base em tais escores. A ordenação é uma tarefa difícil e controversa, especialmente quando se trata de fotografias obtidas por fotógrafos amadores, os quais tipicamente não possuem conhecimento técnico para tal tarefa.

Esta seção inclui uma discussão das abordagens destinadas à extração de elementos de interesse que podem ser importantes para um sistema de análise de fotografias.

Nos algoritmos de extração de informação, o objetivo não é ordenar fotografias, mas extrair elementos de interesse que possam ser utilizados para a sua ordenação, a saber: (1) uma face; (2) um alvo humano; (3) regiões com características indesejáveis (e.g., linhas de dissecação (SHEN et al., 2009)); (4) regiões desfocadas (TONG et al., 2004a); (5) pôr-do-sol (BOUTELL; LUO; GRAY, 2003), dentre outros. A revisão bibliográfica realizada abrangeu abordagens para a detecção de faces e de alvos humanos, a avaliação da paisagem (e.g., detecção de rotação na linha do horizonte), a identificação da classe da imagem (e.g., se é uma fotografia ou uma imagem gráfica), dentre outros.

Por fim, os algoritmos destinados ao agrupamento foram desenvolvidos com o intuito de, a partir de uma função de avaliação, encontrar grupos cuja similaridade entre os componentes é alta.

Nesta pesquisa, há um interesse particular no aprofundamento da investigação das abordagens sobre a avaliação da qualidade das fotografias. Assim, as abordagens sobre extração de informação e agrupamento são brevemente tratadas nas seções 2.1.6 e 2.1.7.

Diversas técnicas fotográficas e regras de qualidade foram definidas por fotógrafos experientes com base em heurísticas. Tais técnicas foram consideradas como as responsáveis pela melhoria da qualidade das fotografias, a ponto de passarem a ser chamadas de regras de composição fotográfica, podendo ser utilizadas para identificar fotografias de maior qualidade, com base nas avaliações de uma ou mais características presentes.

A composição fotográfica, de acordo com Savakis, Etz e Loui (2000), é o fator mais determinante na avaliação positiva da qualidade de uma fotografia pelos consumidores.

A aplicação das regras de composição fotográfica não necessariamente garante melhores resultados estéticos, pois outros fatores são levados em consideração, tais como o alvo. Entretanto, tais regras podem tornar fotografias mais atrativas a observadores humanos (SAVAKIS; ETZ; LOUI, 2000; CAVALCANTI et al., 2010). Paradoxalmente, não é necessário, para uma fotografia parecer mais atrativa, que tenham sido usadas regras de composição. Esta contradição pode ser explicada, em parte, pela existência de outros fatores determinantes da atratividade (e.g., os alvos humanos presentes, sua fotogenia e o local onde a foto foi obtida).

Algumas das regras de composição fotográficas foram posteriormente explicadas por teorias da percepção. A regra-dos-terços (FREEMAN, 2004; BUSSELLE, 1999; GRILL; SCANLON, 1990; HEDGECOE, 2003, 2005) é um bom exemplo: sabe-se que, quando o alvo da fotografia é posicionado nas intersecções das linhas que dividem a imagem em terços, o observador é estimulado, devido à natureza do sistema visual humano, a perceber outras regiões de interesse na fotografia. Outras regras não são tão bem definidas na literatura sobre fotografia ou são definidas em termos de conceitos subjetivos, e.g., a regra que sugere que deve existir um contato visual “sincero” entre os alvos humanos (BUSSELLE, 1999).

As regras de composição fotográfica vêm sendo utilizadas em diversas pesquisas relativas à ordenação de fotografias pela importância (LIU et al., 2010; CAVALCANTI et al., 2010, 2006; DATTA et al., 2006; BANERJEE; EVANS, 2004; BYERS et al., 2004; KE; TANG; JING, 2006; SHEN et al., 2009; LUO; TANG, 2008), a partir das quais identificam-se relações entre algumas regras predefinidas e o julgamento humano. Regras podem ser derivadas de teorias sobre o sistema visual humano, como também por intermédio da experiência de fotógrafos profissionais.

A regra-dos-terços é uma das regras de composição mais exploradas na literatura (LIU et al., 2010; CAVALCANTI et al., 2010, 2006; DATTA et al., 2006; BANERJEE; EVANS, 2004; BYERS et al., 2004). Uma das principais razões é a facilidade de tradução da regra em um algoritmo. Conforme descrita anteriormente, a regra-dos-terços indica que o alvo da fotografia deve, preferencialmente, ser posicionado em um

terço da largura ou altura da imagem (a depender da orientação de tal imagem). As pesquisas existentes divergem sobre como o alvo da fotografia é localizado, i.e., sugere-se a utilização de detectores de faces (CAVALCANTI et al., 2010, 2006; BYERS et al., 2004); informação de baixo nível, tais como bordas e regiões identificadas por meio de atenção visual (ITTI; REES; TSOTSOS, 2005); algoritmos de segmentação, tal como o *mean shift* (LIU et al., 2010; BANERJEE; EVANS, 2004) ou um processo de avaliação da diferença dos *pixels* posicionados em áreas de interesse (DATTA et al., 2006).

Outras regras de composição também são exploradas, entretanto, com menor consenso entre os autores. A regra do *zoom*, por exemplo, possibilita classificar as fotografias de acordo com a distância entre a câmera e o alvo. Distâncias muito grandes ou muito pequenas são penalizadas pelo algoritmo. Uma vez que a detecção do alvo é requerida para tal tipo de análise, Cavalcanti et al. (2006) e Byers et al. (2004) empregaram os resultados fornecidos por detectores de face como principal informação para identificar a posição do alvo. Apesar de detectores de face falharem em algumas ocasiões (e.g., quando os alvos estão de costas), tal abordagem é suficiente para poses fotográficas corriqueiras.

A regra da integridade foi proposta para identificar recortes indesejados no alvo da fotografia. O principal viés desta regra é o alto custo computacional do processo de detecção precisa do alvo em uma fotografia. O uso de razões antropométricas se mostrou efetivo para alvos em poses frontal e vertical. Partindo de alguma informação confiável, e.g., as coordenadas e dimensões de uma face detectada (SHEN et al., 2009; CAVALCANTI et al., 2006), é possível inferir a posição do resto do corpo do alvo, tornando, assim, possível a avaliação de eventuais recortes.

Tanto a regra do *zoom* como a regra da integridade foram desenvolvidas considerando-se que existem informações confiáveis de alto nível, tais como as coordenadas das faces. Isto é uma desvantagem, dado que uma detecção imprecisa pode levar a conclusões inexatas sobre a cena. Existem abordagens que se utilizam de informações de baixo nível (e.g., a intensidade dos *pixels*), ao invés de informação de alto nível. Uma desvantagem é que tais intensidades são calculadas a partir dos canais de cor separadamente, resultando em uma informação possivelmente redundante. Por exemplo, na pesquisa de Datta et al. (2006), algumas regras de composição foram aplicadas

separadamente nos três canais de cor da imagem, alcançando-se 56 regras de composição possivelmente redundantes. A comprovação da redundância foi feita quando tal quantidade de regras foi reduzida para 15, após a poda no resultado do treinamento das máquinas de vetor de suporte (*Support Vector Machines* - SVM) (CORTES; VAPNIK, 1995).

A ordenação de fotografias por importância pode também ser entendida como a seleção de fotografias contendo alvos de interesse e descarte de imagens que apresentem características indesejáveis. Neste sentido, um algoritmo reconhecedor de faces pode ser utilizado como um algoritmo de ordenação por importância, uma vez que pode ser utilizado para selecionar ou descartar fotografias que contenham uma dada face, conhecida ou não.

O reconhecimento de faces pode também ser utilizado para identificar as fotografias que contenham um indivíduo específico e encontrar algum relacionamento entre as imagens devido à presença de um ou vários alvos humanos (CARDOSO; GOMES, 2007). Em um cenário de seleção de fotografias, pode ser muito importante encontrar rapidamente alvos humanos que possam estar presentes em algum conjunto de imagens e analisar o relacionamento existente entre tais alvos (LOUI et al., 2008). O reconhecimento de faces depende fortemente da etapa de detecção de faces. Portanto, imprecisões na detecção de faces podem comprometer a etapa de reconhecimento. Existem, entretanto, abordagens para o reconhecimento de faces robustas ao desalinhamento (YAN et al., 2010).

Detalhes faciais (e.g., marcas de nascença (PIERRARD; VETTER, 2007)) e roupas (GALLAGHER; CHEN, 2008), são também utilizados para melhorar o reconhecimento da face. Campos aleatórios de Markov são utilizados para reconhecer alvos humanos com base em evidências conceituais, e.g., a roupa utilizada (ANGUELOV et al., 2007). O gênero também pode ser uma indicação para o reconhecimento facial e pode ser obtido pelo uso de modelos espaciais de mistura Gaussiana (*spatial Gaussian mixture models* - SGMM) (LI; ZHOU; HUANG, 2009).

Uma vez que as características de baixo-nível da imagem são consideradas pelo consumidor para determinar se uma dada fotografia é melhor do que outra (SERRANO; SAVAKIS; LUO, 2002), detectar automaticamente a presença de tais características

pode ser útil na ordenação das fotografias por importância. Uma destas características é o desfoque (KE; TANG; JING, 2006; LUO; TANG, 2008). Fotografias desfocadas podem ser identificadas a partir de seu gradiente e da análise do espectro do sinal (LIU; LI; JIA, 2008a). A análise espectral do gradiente da imagem também é realizada para identificar núcleos de desfoque em imagens (JI; LIU, 2008).

A análise de alto nível também pode ser empregada para a ordenação das fotografias. A análise estética, por exemplo, tem como objetivo aprender sobre a classificação humana de uma dada fotografia como “boa” ou “ruim”, de acordo com aspectos tanto objetivos como subjetivos. Apesar da dificuldade intrínseca, existem estudos para a compreensão das emoções que as obras de arte evocam nos seres humanos que as observam (YANULEVSKAYA et al., 2008). Os critérios podem ser diversos, e.g., o tempo que um observador dispende observando uma imagem (ENGELKE; MAEDER; ZEPERNICK, 2009). Informações de baixo nível, tais como iluminação, cor (LUO; TANG, 2008), bordas e níveis de brilho (FEDOROVSKAYA; NEUSTAEDTER; HAO, 2008), foram utilizadas para julgar a harmonia de uma fotografia digital (e, eventualmente, de quadros de um vídeo), dado que a harmonia também é um fator que pode ser utilizado por humanos para qualificar uma fotografia.

Além de todos os fatores previamente apresentados, existem aqueles que podem influenciar o julgamento humano. Apesar de que não foram encontradas pesquisas, no âmbito desta investigação, que corroborem tal afirmação, alguns aspectos adicionais devem ser considerados, a saber:

- **Alvos humanos presentes.** Fotógrafos amadores podem considerar uma foto mais ou menos atrativa, dependendo dos alvos humanos que estiverem na fotografia, e.g., uma fotografia com má composição e iluminação pode ser considerada boa se contiver alvos humanos pelos quais o fotógrafo amador possui afeição (e.g., um filho ou uma celebridade). A recíproca também é verdadeira: uma fotografia com boa composição pode ser descartada do conjunto de fotografias de um usuário, caso o alvo não lhe seja conhecido ou querido.
- **Local onde a fotografia foi obtida.** Fotógrafos amadores obtêm grande parte de suas fotografias em viagens. Sendo assim, uma cidade, escolhida para a viagem

de férias, pode parecer mais interessante que a cidade onde o fotógrafo reside.

- **Fotogenia.** A fotogenia é um conceito subjetivo que indica o quanto um rosto humano pode ser considerado esteticamente agradável. Fotografias cujos alvos humanos sejam fotogênicos tendem a ser consideradas superiores.
- **Outras preferências individuais.** Este aspecto engloba todas as demais preferências individuais do usuário, tais como a preferência por fotografias que não obedeçam a regras de composição ou a preferência por determinada categoria de fotografias (e.g., paisagens e natureza).

Apesar dos aspectos previamente mencionados, a ordenação de fotografias pode ser útil para ajudar fotógrafos amadores a identificarem, em um grupo de fotografias pré-selecionadas, aquelas que se destacam.

As regras de composição fotográfica servem, principalmente, para possibilitar a avaliação do posicionamento e relacionamento de um elemento ou grupo de elementos em uma fotografia. Em particular, deseja-se analisar o posicionamento do alvo da fotografia, de modo a enfatizá-lo (GRILL; SCANLON, 1990). Tais regras são comumente adotadas por grande parte dos fotógrafos, tanto amadores quanto profissionais, podendo ser facilmente encontradas na literatura especializada sobre fotocomposição (HURTER, 2004; HEDGECOE, 2005). Toda fotografia tem um alvo ou tema, que é “o que” o fotógrafo desejou capturar no momento em que produziu a fotografia. A análise da fotografia, portanto, visa a avaliar o relacionamento entre o tema da fotografia e o restante do seu conteúdo, de modo a verificar a consistência de tal relacionamento.

Qualquer objeto pode ser o alvo de uma fotografia. Além de elementos concretos, elementos subjetivos também podem ser alvos de fotografias, e.g., o amor ou a solidão. O escopo desta pesquisa restringiu-se a fotografias nas quais os alvos são humanos. Apesar de tal redução de escopo, a maior parte dos métodos apresentados neste capítulo também podem ser aplicados a alvos não-humanos, desde que estes sejam concretos, únicos e bem definidos, ou seja, desde que não exista outro alvo na cena que possa ser confundido com o alvo da fotografia.

Nesta seção, serão apresentadas abordagens para a análise da composição de fotografias. De acordo com Savakis, Etz e Loui (2000), a composição fotográfica é conside-

rada por observadores como o principal fator para o julgamento de uma fotografia. Um dos possíveis motivos pelo qual a composição tem tanta importância é a forma como o cérebro humano interpreta padrões visuais. A teoria de gestalt (KOFFKA, 1955) busca explicar como se dá tal interpretação. Técnicas de composição de fotografias buscam, portanto, canalizar a atenção do observador para pontos de interesse na fotografia.

O objetivo desta seção é discutir abordagens para a análise de fotografias, de modo a avaliá-las objetivamente, levando-se em consideração um conjunto de regras de composição pré-definidas. Regras de composição fotográfica não devem ser encaradas como princípios rígidos, mas, como guias. Vale ressaltar que as regras de composição fotográfica não são os únicos fatores considerados na avaliação de fotografias. Assim, o cumprimento das regras não implica a obtenção de uma boa fotografia. Entretanto, entende-se que, em geral, fotografias obtidas em observância às regras de composição serão melhor avaliadas por observadores humanos do que fotografias que não seguiram ou que descumpriram tais regras.

Existem diversas regras de composição fotográfica. Entretanto, nem toda regra de composição pode ser facilmente traduzida em um algoritmo. Quatro são as regras de composição transformadas em algoritmos e discutidas neste documento: três advindas de esforços anteriores - a regra dos terços, a regra do zoom e a regra da integridade (CAVALCANTI et al., 2006; CAVALCANTI, 2007) - e uma quarta regra - regra do espaço negativo - proposta nesta pesquisa. Assumindo que apenas fotografias contendo alvos humanos são consideradas, as informações necessárias por tais algoritmos para a análise da composição fotográfica podem ser derivadas da posição e dimensões da face, as quais podem ser obtidas a partir de detectores clássicos de faces e de olhos (ROWLEY; BALUJA; KANADE, 1998a, 1998b; BRADSKI, 2000; LEITE et al., 2007; FASEL; FORTENBERRY; MOVELLAN, 2005).

2.1.1 Antropometria

O método proposto por Cavalcanti (2007) utiliza as coordenadas das faces detectadas para inferir o posicionamento dos alvos humanos cujas faces foram detectadas. Para que tal inferência possa ser realizada, relações antropométricas são utilizadas.

As relações antropométricas, são relações entre as medidas das partes do corpo hu-

mano. Tais relações são verificadas como verdadeiras para grande parcela da população, sendo frequentemente utilizadas em algumas áreas de conhecimento (e.g., ergonomia, arquitetura, anatomia e artes plásticas) para otimizar as dimensões de produtos (TILLEY, 2001; LIDWELL; HOLDEN; BUTLER, 2003).

Para alguns grupos em particular, as relações antropométricas apresentam diferenças, e.g., as cabeças das crianças no momento do nascimento têm a dimensão aproximada do restante do corpo (TILLEY, 2001). Relações utilizadas nesta pesquisa, contudo, são válidas para corpos de indivíduos adultos. Eis algumas relações típicas (TILLEY, 2001; LIDWELL; HOLDEN; BUTLER, 2003):

$$h_{\text{corpo}} = 7 \times h_{\text{cabeça}}, \quad (2.1)$$

$$h_{\text{cintura}} = 4 \times h_{\text{cabeça}}, \quad (2.2)$$

$$w_{\text{ombros}} = 3 \times w_{\text{cabeça}}, \quad (2.3)$$

sendo $h_{\text{cabeça}}$ e $w_{\text{cabeça}}$ a altura e largura da cabeça de um indivíduo, h_{corpo} a altura do corpo, h_{cintura} a distância entre o topo da cabeça e a linha da cintura e w_{ombros} a largura dos ombros. Vale ressaltar que estas relações são válidas para indivíduos adultos.

2.1.2 Regra dos terços

A regra dos terços é, provavelmente, a mais conhecida das regras de composição fotográfica, sendo encontrados guias visuais para esta regra nos visores de um número significativo de modelos de câmeras digitais atuais. A regra sugere que o alvo da fotografia deve estar em um dos terços (horizontal ou vertical) da área do enquadramento.

A proporção dos terços é adotada quando se imagina duas linhas horizontais e duas linhas verticais dividindo a área de enquadramento em 3 partes iguais, tanto na horizontal quanto na vertical. A composição dos elementos visuais próximos dessas linhas imaginárias ou de seus pontos de interseção produz o que se denomina *balanço dinâmico*, uma vez que, ao invés do balanço estático, monótono e centralizado, a composição fora de centro produz um efeito visual balanceado e harmônico.

Assim, as coordenadas y das linhas de terço horizontais são iguais a $\frac{w}{3}$ e $\frac{2 \times w}{3}$ e,

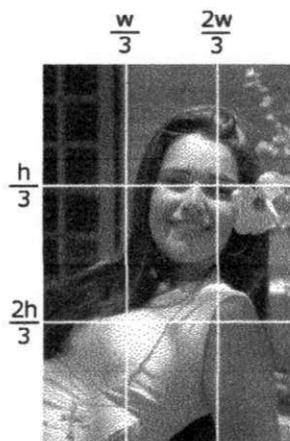
analogamente, as coordenadas x das linhas de terço verticais são $\frac{h}{3}$ e $\frac{2 \times h}{3}$, sendo w e h a largura e a altura da fotografia, respectivamente.

A aplicação da regra dos terços, de acordo com fotógrafos profissionais, evita a centralização do alvo da fotografia, o que pode resultar em uma fotografia estática. Uma fotografia é denominada estática quando não possui elementos visuais suficientes para conduzir a atenção do observador ao longo de suas dimensões, o que faz com que o cérebro considere-a não atrativa (HURTER, 2004; HEDGECOE, 2005).

Em retratos³, o centro de interesse mais adotado são os olhos, os quais são idealmente posicionados no ponto dos terços (interseção de linhas de terço verticais e horizontais).

Na Figura 2.1, é mostrado um retrato no qual os olhos da modelo foram corretamente posicionados, de acordo com a regra dos terços. Dependendo da distância ao alvo humano, poderia-se ter posicionado o ponto dos terços em qualquer outra posição da cabeça da modelo. Também são indicadas as posições dos terços horizontais e verticais.

Figura 2.1: Regra dos terços: um dos olhos está sobre um dos quatro pontos dos terços, enquanto o outro está sobre a linha dos terços horizontal superior.



Para avaliar a conformidade à regra dos terços, Cavalcanti et al. (2010) calcularam as oito distâncias euclidianas entre os olhos, detectados pelo algoritmo de Leite et al. (2007), e os quatro pontos dos terços. A menor de tais distâncias s é normalizada,

³Termo utilizado para denotar fotografias cujo alvo são humanos.

conforme explicitado na Equação 2.4, no intervalo $[-1,1]$.

$$n = 1 - \left(2 \times \frac{s}{m}\right) \quad (2.4)$$

sendo m a distância máxima teórica (com o centro de interesse posicionado num dos cantos da imagem), calculada como mostrado na Equação 2.5.

$$m = \frac{\sqrt{w^2 + h^2}}{3} \quad (2.5)$$

A condição $n = 1$ implica que, na fotografia, houve uma coincidência entre os olhos do alvo e o ponto dos terços. À medida que a distancia entre os olhos e tais pontos aumenta, o valor de n diminui até o valor mínimo ($n = -1$), representando o pior posicionamento possível, com relação à regra dos terços.

2.1.3 Regra do *zoom*

A regra do *zoom* possibilita a análise das distâncias relativas entre os alvos e a câmera. No caso de alvos humanos, uma estratégia é avaliar as dimensões do corpo (ou das partes visíveis do corpo), com relação às dimensões da fotografia.

Além das dimensões do alvo, a regra do *zoom* também permite a identificação do espaço existente entre o alvo e as bordas da fotografia (HURTER, 2004). Tal espaço é comumente denominado *room*. Similarmente, *headroom* é o espaço existente entre a cabeça e as bordas da fotografia (BUSSELLE, 1999).

Fotografias contendo cabeças posicionadas próximas às bordas dão a impressão de que as cabeças dos alvos foram forçadas contra as bordas. Por outro lado, se as cabeças presentes na fotografia estiverem longe da borda superior, porém extremamente próximas à borda inferior, dão a impressão de decapitação do(s) alvo(s). Ambos são efeitos artisticamente indesejáveis em composições fotográficas rotineiras (HURTER, 2004).

Algumas heurísticas são definidas por fotógrafos profissionais, com o objetivo de decidir quão perto do alvo a câmera deve estar (BUSSELLE, 1999). Contudo, assim como acontece com a maioria das outras regras de composição, essas heurísticas não

são consensuais. A partir da análise de algumas pesquisas nas áreas de fotografia e de cinema (HEDGECOE, 2005; HURTER, 2004; TREMBLAY, 2003; OKAZAKI, 1998), Cavalcanti (2007) definiu quatro padrões de distância para o tema da fotografia:

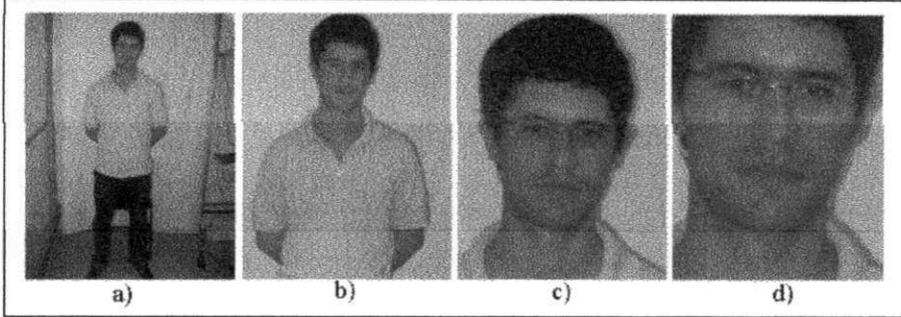
- *Close-up* extremo (*extreme close-up*): quando praticamente toda a área da fotografia é ocupada apenas pela face do alvo humano. É preciso que haja o recorte de alguma parte da cabeça do alvo sem, contudo, haver recorte da boca ou dos olhos;
- *Close-up*: quando a região do topo da cabeça até os ombros do alvo humano é visível na fotografia. Do topo da cabeça à parte superior da fotografia, não deve existir uma distância maior que a altura de meia cabeça do alvo;
- Média distância (*medium-shot*): quando a região do topo da cabeça à linha da cintura do alvo humano é visível na fotografia. Do topo da cabeça à parte superior da fotografia, não deve existir uma distância maior que a altura de uma cabeça do alvo; e
- Longa distância (*long-shot*): quando todo o corpo do alvo humano é visível. Não deve existir uma distância maior que a altura de uma cabeça da cabeça do alvo à borda superior nem dos pés do alvo à borda inferior da fotografia;

Portanto, uma forma de analisar se a distância do alvo às bordas está adequada é verificar se algum dos padrões de distância anteriormente mencionados foi adotado. Para tanto, pode-se identificar o posicionamento do alvo na fotografia.

De modo a simplificar e agilizar este processo, Cavalcanti (2007) desenvolveu uma abordagem que utiliza medidas antropométricas (descritas na Seção 2.1.1) para estimar as dimensões e posicionamento do corpo de um alvo humano por meio das coordenadas da face e efetuar o recorte da fotografia. Nesta pesquisa, foi feita uma adaptação da abordagem de Cavalcanti (2007) para, ao invés de se modificar a fotografia, atribuir-lhe um escore proporcional à conformidade à regra.

Os padrões de *zoom* supramencionados, i.e., longa distância, média distância, *close-up* e *close-up* extremo, estão ilustrados na Figura 2.2.

Figura 2.2: Ilustração dos níveis de *zoom*: (a) longa distância; (b) média distância; (c) *close-up*; (d) *close-up* extremo.



Os limiares para o cálculo da distância do tema (no caso, a face do alvo humano) às bordas da fotografia foram definidos empiricamente, de forma a respeitar o *headroom* e as definições dos padrões supramencionados. As proporções utilizadas são mostradas na Tabela 2.1. Os valores indicam as distâncias até as bordas com relação à altura (para as bordas superior e inferior) e largura (bordas laterais) da face, ou seja, para uma fotografia que tem por meta um *close-up*, da lateral da face até a borda esquerda deve haver pelo menos $0,8 \times w_{face}$ pixels, sendo w_{face} a largura da face, também expressa em pixels.

Tabela 2.1: Parâmetros para a regra do *zoom*.

Padrão de zoom / Posicionamento	superior	inferior	esquerda	direita
<i>Longa distância</i>	1,0	8,0	1,0	1,0
<i>Média distância</i>	0,7	3,0	0,8	0,8
<i>Close-up</i>	0,6	0,6	0,4	0,4
<i>Close-up extremo</i>	0,08	0,6	0,3	0,3

Assim, calcula-se a distância para cada borda da fotografia (superior, inferior, esquerda e direita). A diferença com relação ao padrão é calculada somando-se os módulos das diferenças entre as distâncias para todas as bordas e as distâncias indicadas por cada padrão, conforme explicitado na Equação 2.6.

$$D_p = \sum_{b=1}^4 M_{p,b} \quad (2.6)$$

em que D_p é a diferença do posicionamento existente na fotografia para um dado padrão de distância p e M é uma matriz bi-dimensional, na qual cada linha representa um padrão de distância (i.e., longa distância, média distância, *close-up* e *close-up* extremo, respectivamente) e cada coluna uma extremidade da fotografia (i.e., superior, inferior, esquerda e direita, respectivamente).

De posse das distâncias, calcula-se qual o padrão de composição mais próximo, P_{prox} , conforme explicitado na Equação 2.7.

$$P_{prox} = \operatorname{argmin}_i D, i \in \{\text{longa dist.}, \text{média dist.}, \text{close-up}, \text{close-up extremo}\} \quad (2.7)$$

O erro absoluto cometido ao se utilizar a Equação 2.7 pode ser definido como a diferença entre a distância calculada e o padrão de distância mais próximo de ser alcançado. Assim, ao se calcular o erro, três casos são possíveis:

1. a distância calculada é maior do que aquela correspondente ao padrão longa distância;
2. a distância calculada é menor do que aquela correspondente ao padrão *close-Up* extremo;
3. a distância calculada é intermediária àquelas correspondentes a dois outros padrões.

Para os casos 1 e 2, definiu-se, empiricamente ao se observar as situações extremas, que se a distância for maior que $10 \times h_{face}$ ou menor que $0,04 \times h_{face}$, respectivamente, o score será -1. Tais limiares foram calculados com base nas definições das regras. No caso do padrão longa distância, deve haver no máximo o espaço de uma cabeça entre a cabeça e a parte superior da imagem e os pés do alvo e a parte inferior da imagem, ou seja, oito faces (a altura ideal da figura humana) mais duas faces de espaço complementar. Por sua vez, o valor referente ao *close-up* extremo foi definido experimentalmente.

Para o terceiro caso, definiu-se empiricamente que o score diminui gradativamente de 1 até -1, à medida que a diferença para o padrão mais próximo se aproxima da metade da distância entre o padrão mais próximo e o segundo padrão mais próximo.

2.1.4 Regra da integridade

A partir desta regra, verificam-se se partes importantes do corpo do alvo humano estão sendo “cortadas” da fotografia. Apesar de algum nível de corte do alvo ser tolerável, é desejável que partes tais como a face, os olhos, as mãos e as articulações sejam preservados ou tal recorte poderá comprometer a qualidade da composição.

Nesta pesquisa, optou-se por analisar a integridade da cabeça, a qual é essencial em qualquer fotografia contendo alvos humanos. As dimensões aproximadas da cabeça são estimadas a partir da distância d_{olhos} entre os olhos, considerando a aplicação de algum detector existente de faces e de olhos, neste caso, utilizando-se a abordagem desenvolvida por Leite et al. (2007).

A região retornada por um detector de faces poderia servir como estimativa para a cabeça. Contudo, percebe-se que há muita instabilidade nas dimensões desta região nos detectores existentes (e.g., no detector proposto por Viola & Jones (VIOLA; JONES, 2001), disponível na biblioteca OpenCV (BRADSKI, 2000)).

Desta forma, para faces frontais, a distância entre os olhos tende a ser mais estável na estimativa da região da cabeça. Cabe observar que a estratégia descrita é, na verdade, uma aproximação de uma estrutura 3D (a cabeça humana), a partir da área de sua projeção 2D frontal. A largura da cabeça ($w_{cabeça}$) é calculada a partir da Equação 2.8.

$$w_{cabeça} = 1.2 \times d_{olhos} \quad (2.8)$$

De forma similar, pode-se calcular a altura da cabeça ($h_{cabeça}$) a partir da Equação 2.9.

$$h_{cabeça} = 1.5 \times w_{cabeça} \quad (2.9)$$

As constantes nas Equações 2.8 e 2.9 foram obtidas a partir dos estudos antropométricos apresentados na Seção 2.1.1.

O conceito de *integridade* pode ser expresso como a razão entre a área recortada da cabeça ($areaRecorte_{cabeça}$) e a área total da cabeça ($areaTotal_{cabeça}$). O cálculo

de $areaRecorte_{cabeça}$ é efetuado a partir da soma dos resultados dos recortes do topo (Equação 2.10) e da lateral (Equação 2.11) da cabeça, conforme expresso na Equação 2.12.

$$recorteTopo_{cabeça} = \begin{cases} w_{cabeça} \times |y_1|, & \text{se } y_1 < 0; \\ w_{cabeça} \times (y_2 - h), & \text{se } y_2 > h; \\ 0, & \text{caso contrário.} \end{cases} \quad (2.10)$$

$$recorteLateral_{cabeça} = \begin{cases} h_{cabeça} \times |x_1|, & \text{se } x_1 < 0; \\ h_{cabeça} \times (x_2 - w), & \text{se } x_2 > w; \\ 0, & \text{caso contrário.} \end{cases} \quad (2.11)$$

$$areaRecorte_{cabeça} = recorteTopo_{cabeça} + recorteLateral_{cabeça} \quad (2.12)$$

Existem dois tipos de violação à regra: o recorte do topo da cabeça (expresso por $recorteTopo_{cabeça}$ na Equação 2.10) e o recorte da lateral da cabeça (expresso por $recorteLateral_{cabeça}$ na Equação 2.11). As coordenadas (x_1, y_1) e (x_2, y_2) , representam os vértices superiores esquerdo e inferior direito da área recortada, respectivamente. Como antes, w e h são a largura e a altura da fotografia, respectivamente. Esta avaliação é possível uma vez que, na ocorrência de recortes da face, a maioria dos detectores de face inferem coordenadas fora dos limites da fotografia (i.e., valores negativos ou coordenadas maiores que a largura e altura da fotografia).

Por sua vez, a área total da cabeça, $areaTotal_{cabeça}$, é definida na Equação 2.13.

$$areaTotal_{cabeça} = w_{cabeça} \times h_{cabeça} \quad (2.13)$$

Por fim, a Equação 2.14, contém a expressão destinada ao cálculo da integridade $integridade_{cabeça}$.

$$integridade_{cabeça} = 1 - \left(2 \times \frac{areaRecorte_{cabeça}}{0,5 \times areaTotal_{cabeça}} \right) \quad (2.14)$$

Os detectores testados (ROWLEY; BALUJA; KANADE, 1998a, 1998b) não definem explicitamente um nível máximo permitido de oclusão. Entretanto, verificou-se

experimentalmente que este nível é próximo de 50% da área total da cabeça. Assim, a regra da integridade é normalizada na faixa de -1 a 1, tendo 50% como valor máximo de recorte (multiplicando-se por 0,5 a área total da cabeça). Na ocorrência de recortes maiores que 50%, o valor da integridade é definido em -1.

Eventualmente, a cabeça é recortada tanto na extremidade superior quanto na extremidade inferior da fotografia. Neste caso, a violação da integridade resulta em y_1 menor que 0 e x_2 maior que a largura da fotografia. Nesta última situação, o valor da largura da cabeça $w_{cabeça}$ precisa ser adicionado ao valor da interseção das áreas recortadas, o qual foi subtraído duas vezes. Neste caso, a área recortada da cabeça é calculada a partir da Equação 2.15.

$$areaRecorte_{cabeça} = recorteTopo_{cabeça} + recorteLateral_{cabeça} - area_{intersecção} \quad (2.15)$$

Na Figura 2.3, é ilustrada uma situação em que tanto o topo quanto as laterais da cabeça foram recortados. Em algumas situações, como naquelas em que se busca um *close-up* extremo, tal tipo de violação é aceitável, sendo preferível o recorte de parte da cabeça do que parte dos ombros, por exemplo. Mas na existência de espaço disponível, como tipicamente ocorre em fotografias de corpo inteiro, tal tipo de recorte é inaceitável.

Figura 2.3: Exemplo de violação da integridade: topo e laterais da cabeça foram cortados.



2.1.5 Análise da harmonia das cores de uma imagem

Uma vez que parte desta pesquisa é relacionada à análise de cores, é importante incluir uma breve explicação sobre o funcionamento do sistema visual humano. Vale frisar

que os conceitos explorados nesta seção dizem respeito ao sistema visual humano sadio. Desta forma, qualquer generalização feita a seguir sobre os seres humanos se aplica tão somente àqueles que não possuem distúrbios no sistema visual.

A estrutura do olho humano diferencia-se das estruturas dos olhos da maioria dos animais, devido à larga gama de cores às quais é sensível, compreendendo um espectro eletromagnético que vai de, aproximadamente, $0,43 \mu m$ (correspondente à cor violeta) até $0,79 \mu m$ (correspondendo à cor vermelha) (GONZALEZ; WOODS, 2001). A retina do olho humano é constituída pela retina pigmentar e pela retina sensorial. A retina sensorial contém camadas de neurônios fotorreceptores, denominados cones e bastonetes. Os cones estão associados ao processo de formação de imagens mais nítidas na fóvea, sendo os principais responsáveis pela identificação das cores. De acordo com a teoria do triestímulo (PRATT, 2007; FAIRCHILD, 2005), há três tipos de cones: (1) aqueles sensíveis à radiação com comprimentos de onda mais elevados, associados à visualização da luz vermelha; (2) aqueles sensíveis à radiação com comprimentos de onda intermediários, associados à visualização da luz verde; e (3) aqueles sensíveis à radiação com comprimentos de onda mais baixos, associados à visualização da luz azul (FAIRCHILD, 2005). Os bastonetes não são tão sensíveis às cores quanto os cones (HOOK, 2004), são mais sensíveis à luminosidade, sendo os principais responsáveis pela visão noturna (razão pela qual, não é possível identificar cores adequadamente em condições de baixa luminosidade) (FAIRCHILD, 2005).

O processo de interpretação das cores é um fenômeno psicofísico que ainda não foi completamente compreendido (GONZALEZ; WOODS, 2001). Entretanto, sabe-se que a percepção da cor é diferente em diferentes indivíduos, apesar de alguns conceitos de percepção, tais como a temperatura da cor (quente, fria), e a tonalidade, dentre outros, serem percebidos de forma similar por seres humanos.

As cores são utilizadas em documentos e ilustrações para transmitir informações adicionais ao conteúdo escrito. Em decorrência da capacidade do ser humano de perceber e interpretar de forma semelhante as tonalidades existentes, as cores são frequentemente utilizadas para guiar a atenção humana (MORTON, 2004). Muitos estudos psicológicos sugerem a associação de certos sentimentos à presença de determinadas

cores, e.g., a transmissão de tranquilidade é conferida à cor verde (PRO, 2008; MORTON, 2008b), enquanto a cor vermelha é associada à sensação de ansiedade, devido ao seu grande comprimento de onda e, acredita-se, por ser a cor do sangue (PRO, 2008; PETERSON, 2009; CHERRY, 2009). Em consequência, o vermelho é utilizado para sinalizar elementos para os quais se deseja atenção especial: a cor do semáforo e das placas de sinalização que exigem mais atenção (e.g., a placa de “PARE” utilizada pelos departamentos de controle de trânsito). Se a mesma placa fosse pintada de cor verde, provavelmente não receberia dos motoristas a mesma atenção (MORTON, 2008a).

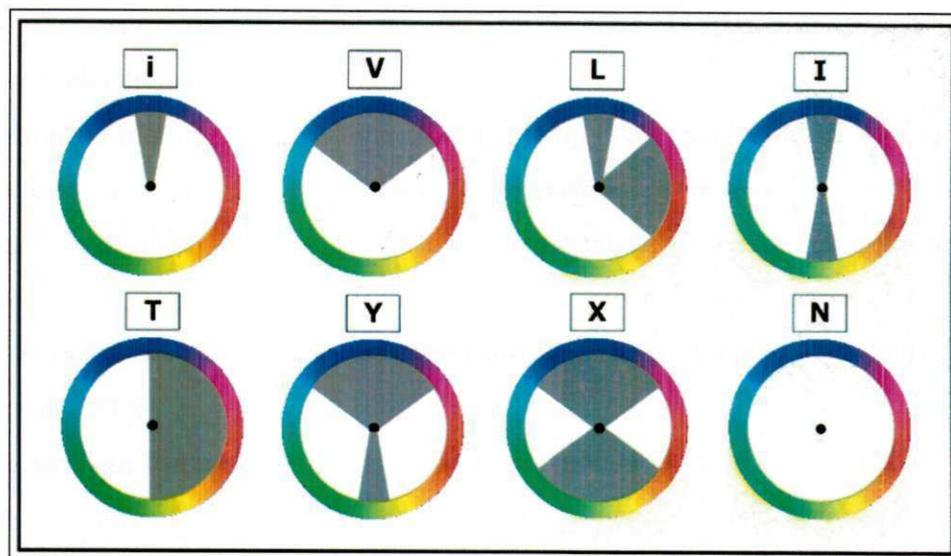
Adicionalmente aos sentimentos despertados pelas cores isoladamente, também são notórias as sensações associadas a combinações específicas de cores (ARNHEIM, 1980; PRO, 2008; GARAU; BRUNO; ARNHEIM, 1993). Um exemplo clássico é a combinação de cores utilizada pela maioria das redes de lanchonetes ocidentais. O vermelho é combinado ao amarelo, o qual, individualmente, transmite a sensação de fome (PRO, 2008) para induzir nos clientes o desejo de comer em demasia e rapidamente. Psicofisicamente, esta é uma combinação de cores que resulta em estímulos luminosos com alto comprimento de onda. Uma vez que o centro da retina é repleto de cones sensíveis a estímulos luminosos com comprimento de onda elevado e intermediários (GONZALEZ; WOODS, 2001; FAIRCHILD, 2005), combinações como esta causam desconforto ao observador, mas cumprem com eficiência o papel de induzir a clientela a comer muito e rapidamente.

Nesta pesquisa, a noção de harmonia de cores está alinhada às pesquisas de Matsuda (1995) e Tokumaru, Muranaka e Imanishi (2002), ambas originadas da pesquisa de Itten (1960), os quais adotaram a codificação de cores proposta por Munsell (MUNSELL, 1905, 1969; FAIRCHILD, 2005). Na pesquisa desenvolvida por Tokumaru, Muranaka e Imanishi (2002), foram propostos oito padrões de distribuições dos níveis de contraste de cores. Uma determinada imagem, cujos níveis de contraste se distribuam similarmente a um destes padrões, possui cores em harmonia.

Neste sentido, Cohen-Or et al. (2006) propuseram uma abordagem para corrigir a distribuição das cores de uma imagem, de maneira que esta fique em conformidade com

uma das paletas padronizadas de cor ilustradas na Figura 2.4.

Figura 2.4: Padrões de distribuição dos níveis de contraste. Fonte: adaptado de Cohen-Or et al. (2006).



Na Figura 2.4, são ilustradas as distribuições das matizes de todos os pontos da imagem. Os padrões foram definidos a partir de ângulos, de modo que uma dada distribuição possa ser utilizada, caso haja uma rotação constante das intensidades dos diferentes níveis de contraste. Por isso, escolheu-se o espaço de cor HSV, no qual o matiz é um parâmetro angular, variante entre 0 e 360 graus. Avalia-se, na abordagem de Cohen-Or et al. (2006), portanto, se a distribuição de cores de uma imagem, comporta-se como um dos 8 padrões ilustrados nesta figura.

Sawant e Mitra (2008) propuseram uma adaptação da abordagem proposta por Cohen-Or et al. (2006), com fins ao processamento de vídeos. Basicamente, a adaptação em questão não apresenta contribuições adicionais para a análise da harmonia das cores, sendo apenas um estudo afim.

No estudo de Obrador (2006), é proposto um sistema semelhante àquele adotado nesta pesquisa, no qual são utilizadas as definições de distribuição de cor descritas por Cohen-Or et al. (2006). Na pesquisa de Hou e Zhang (2007), por outro lado, é proposto um método para a harmonização de cores inspirado em conceitos de cor obtidos mediante o uso de protótipos de cor. Para a obtenção de uma imagem com

No topo da Figura 2.5, é mostrado o círculo cromático. Na primeira fileira, são exibidos os padrões análogo e complementar. Na segunda fileira, são exibidos os padrões triádico e complementar dividido. Por fim, na terceira fileira, são exibidos o padrão quadrado e o padrão tetrádico ou duplo complementar. Inicialmente, divide-se o círculo em doze partes (consequência da divisão do círculo em três cores primárias, as quais, duas a duas, produzem mais três cores secundárias que por sua vez, se combinadas às cores primárias dão origem a mais 6 cores terciárias).

Tais padrões podem ser utilizados para, dada uma cor, simplificar a escolha das demais cores componentes do esquema de cor a ser utilizado. Dada uma cor do círculo cromático e tomando-se como exemplo o padrão análogo, devem ser escolhidas as duas cores adjacentes para complementar o esquema. Os esquemas de cor são também inspirados na ideia de avaliar a distribuição de cores ao longo do círculo de matizes.

Os padrões exemplificados na Figura 2.5 são relativos a uma cor principal, no caso, a cor verde. Para se obter o mesmo padrão, a partir de outra cor principal, basta centralizar os padrões, tomando-se como referência a cor desejada e mantendo-se as demais relações.

Material adicional sobre cores podem ser encontrados nos estudos de Kobayashi (1987, 1990). Exemplos de esquemas de cor podem ser encontrados *on-line* (COLOR, 2009).

Nas abordagens previamente descritas, o principal objetivo é modificar as cores das fotografias, até que um determinado padrão de distribuição de cores seja obtido. Entretanto, pode-se destacar como um contraexemplo a pesquisa de Yao et al. (2011). Neste caso, foi feita uma análise da qualidade da fotografia com base na harmonia das cores. Em tal pesquisa, a análise das cores foi feita em todas as cores da imagem. Aplicou-se o algoritmo *k-means* (BISHOP, 1995), a partir de um número significativo de núcleos, sendo feitas combinações três a três desses *k* núcleos. Por fim, tais combinações foram comparadas a combinações obtidas em uma base de fotografias classificadas como boas.

A abordagem supracitada apresenta duas grandes limitações. A primeira é o fato da análise de cores requerer a existência de uma base de fotografias previamente rotuladas significativamente volumosa. Além disto, a coincidência de cores com aquelas obtidas

em uma base de fotografias não garante uma combinação de cores apropriada. Pode acontecer de uma combinação de cores ser apropriada para um certo tipo de alvo, mas não o ser para outro. A segunda limitação relaciona-se à forma com que as cores da imagem são obtidas para a comparação. Na abordagem de Yao et al. (2011), todos os *pixels* da imagem foram considerados. Assim, as regiões não foram avaliadas individualmente. Conseqüentemente, regiões importantes da imagem podiam ser ignoradas na formação de trios de cores, caso não apresentassem uma quantidade significativa de *pixels* com aquela cor.

A última abordagem incluída nesta subseção, igualmente relacionada à análise da harmonia cromática de uma imagem, foi proposta por Sauvaget et al. (2010) e se inspira na análise realizada por Itten (1961). A abordagem proposta por Sauvaget et al. (2010) possui vantagens, com relação à pesquisa de Cohen-Or et al. (2006), por não depender de um padrão pré-estabelecido de distribuição de cores e se inspirar em uma abordagem de análise da harmonia cromática já consolidada (ITTEN, 1961).

A abordagem proposta por Sauvaget et al. (2010) tem como objetivo a coloração de imagens monocromáticas. Assim, a partir das cores previamente escolhidas pelo usuário para compor uma dada imagem, calcula-se, automaticamente, qual é a cor que deveria ser utilizada e em qual quantidade, de modo a manter a harmonia entre as cores existentes.

A harmonia, ao qual se refere o parágrafo anterior, utiliza-se da tabela de proporções de cores definida por Itten (1961). Para cada cor catalogada nesta tabela, há a proporção em que as demais cores igualmente presentes na imagem devem ser utilizadas, de modo que a harmonia cromática entre elas seja mantida. Os valores de proporção definidos por Itten (1961) são apresentados na Tabela 2.2.

Tabela 2.2: Proporções de cor definidas por Itten (1961).

Vermelho	Laranja	Amarelo	Verde	Azul	Violeta
60	40	30	60	80	90

Os valores anteriormente apresentados dividem o círculo cromático nos ângulos indicados pela proporção de cor. Assim, todas as cores que são representadas com um valor de matiz entre 0 e 59 são consideradas vermelhas, todas as que estiverem entre 60

e 99 são consideradas da cor laranja e assim por diante. A partir desta tabela, calcula-se a proporção de uma cor em função das demais cores presentes. Sendo assim, para se calcular, por exemplo, as proporções cromáticas ideais em uma imagem que dispõe das cores amarela, azul e verde, deve-se calcular a proporção pelo quociente entre a proporção da cor e o somatório das demais cores presentes. Deste modo, espera-se que sejam mantidas para as cores amarela, azul e verde as proporções de $\frac{30}{30+80+60}$, $\frac{80}{30+80+60}$ e $\frac{60}{30+80+60}$, respectivamente.

A principal vantagem desta abordagem em relação à abordagem proposta por Cohen-Or et al. (2006), é que há uma relação entre a quantidade de cor utilizada e a harmonia, enquanto a segunda abordagem possui maior foco na adequação aos padrões de cor previamente definidos.

Os estudos apresentados evidenciam o crescente interesse no tema em questão. Ainda assim, há muito espaço para o aprimoramento da modelagem e análise cromática, e.g., a análise da harmonia das cores a partir dos principais alvos presentes na imagem.

2.1.6 Extração de informações

Esta seção inclui uma discussão das abordagens destinadas à extração de elementos de interesse que podem ser importantes para um sistema de análise de fotografias. A revisão bibliográfica efetuada abrangeu abordagens para a detecção de faces e alvos humanos, a avaliação da paisagem (e.g., detecção de rotação na linha do horizonte) e a identificação da categoria a que pertence uma imagem (e.g., se é uma fotografia ou uma imagem gráfica), dentre outros tópicos.

Via de regra, o propósito dos estudos revisados não é ordenar imagens, mas extrair informação de imagens. Tais informações podem ser utilizadas como sendo fontes auxiliares de informação para a ordenação de fotografias. Elementos de interesse podem ser quaisquer elementos que o usuário deseje procurar: (1) uma face; (2) um alvo humano; (3) regiões com características indesejáveis, e.g., linhas de dissecação (SHEN et al., 2009); (4) regiões desfocadas (TONG et al., 2004a; LIU; LI; JIA, 2008b); (5) sol poente (BOUTELL; LUO; GRAY, 2003), dentre outros.

Geralmente, a informação extraída pelas diferentes abordagens envolvem a construção de um modelo de classificação do elemento de interesse (e.g., a partir de um

processo de aprendizagem estatística). Árvores de decisão são tipicamente empregadas para identificar corretamente classes que possuem um número reduzido de restrições, tanto numéricas quanto categóricas (e.g., cores e número de indivíduos). O algoritmo de classificação ID3 (QUINLAN, 1986) foi utilizado para diferenciar fotografias de imagens vetoriais (e.g., um logotipo, um desenho ou outras imagens artificialmente geradas). A árvore de decisão foi treinada com 1200 imagens. Foi obtida uma taxa de acerto de 95,6% na distinção das classes. Este resultado foi verificado a partir de validação cruzada com 10-*folds*.

Máquinas de vetor de suporte (*Support Vector Machines* - SVM) (CORTES; VAPNIK, 1995) são amplamente utilizadas para a detecção de características (SERRANO; SAVAKIS; LUO, 2002; BOUTELL; LUO; GRAY, 2003; TONG et al., 2004b; HAN; AWAD; SUTHERLAND, 2009). Quando existe uma grande quantidade de dados e uma grande quantidade de componentes em um modelo SVM treinado, a alta correlação que pode existir entre os componentes pode prejudicar a classificação. Portanto, também se emprega a análise dos componentes principais (*principal component analysis* - PCA) (SHLENS, 2005) para reduzir a dimensionalidade do espaço de busca (TONG et al., 2004b).

Alguns métodos para a extração de informação são destinados à detecção de informação relacionada a humanos, tais como face, olhos, pele, pose, etc. Uma vez que grande parte das fotografias tem como alvo principal o ser humano, tais métodos são importantes para um sistema de análise de fotografias.

As pesquisas em detecção de faces têm como foco detectores invariantes a rotação, multi-visões e escala. Características discriminantes (WANG; JI, 2007), características de baixo-nível (DESTRERO et al., 2009), operações morfológicas e limiarização (TSAO et al., 2010), também podem ser utilizadas com tal objetivo.

A presença na imagem de tons similares ao da pele humana também é uma evidência importante da presença de alvos humanos na fotografia. Diversas abordagens foram propostas, as quais são passíveis de categorização em função da informação utilizada para a identificação da tonalidade: um único *pixel* ou uma região (KRUPPA; BAUER; SCHIELE, 2002). A classificação da pele pode ser realizada a partir de SVM e de métodos para a segmentação de regiões (HAN; AWAD; SUTHERLAND, 2009).

Enquanto a evidência da presença de alvos humanos pode ser obtida por intermédio de detectores de pele e face, existem abordagens que detectam alvos humanos diretamente nas imagens. Abordagens recentes utilizam padrões binários locais (*local binary patterns* - LBP) para a detecção de alvos humanos via duas variantes: LBP-Semântico e LBP-Fourier (MU et al., 2008). A detecção de alvos humanos pode ser feita a partir da adoção de regras temporais nebulosas quantificadas (*quantified fuzzy temporal rules*) para a representação do conhecimento de dados espaciais relativos aos alvos humanos, os quais são aprendidos por intermédio de abordagens evolucionárias (MUCIENTES; BUGARÍN, 2010).

Alguns estudos têm sido desenvolvidos para a detecção de alvos humanos em contextos específicos, e.g., a detecção de pedestres (PAISITKRIANGKRAI; SHEN; ZHANG, 2008) e a identificação de atores em atividades como 'luta' ou 'assalto' (RYOO; AGGARWAL, 2009).

Além disto, a pose de alvos humanos também pode ser utilizada no processo de análise de fotografias. As abordagens mais populares fazem uso das restrições de posicionamento e rotação que as partes do corpo humano impõem, entre si, e.g., a cabeça é diretamente conectada aos ombros, não sendo possível encontrá-la sob os pés. Assim, se uma cabeça é encontrada, deve-se buscar os ombros nas proximidades. Existem abordagens que utilizam restrições nas variações das articulações a partir do uso de SVM (MAIK et al., 2010) e características Haar, para vídeos (BISSACCO; YANG; SOATTO, 2007).

Apesar dos alvos humanos serem, provavelmente, a categoria mais fotografada, outros tipos de alvos são possíveis, tais como objetos, animais, etc. Por muitas vezes, os objetos são o principal alvo da fotografia, mas, em outras ocasiões, apenas interagem com o alvo humano, resultando em uma fotografia de composição mais complexa. Estudos foram desenvolvidos para investigar a detecção de objetos com base na forma (ROUSSON; PARAGIOS, 2008; GORELICK; BASRI, 2009). Também podem ser empregados, para uma abordagem mais genérica, o algoritmo *Scale-Invariant Feature Transform* (SIFT) (LEE et al., 2010) e a análise das variações de tonalidade da imagem (STEIN; STEPLETON; HEBERT, 2008).

Ao invés de detectar um tipo específico de objeto, também se pode identificar as

regiões da imagem que possuam alguma correspondência. Neste sentido, algoritmos destinados à segmentação de imagens são de fundamental importância na análise das fotografias. Existem diversas abordagens para a segmentação de imagens.

Dentre as técnicas utilizadas para a segmentação de imagens, podem-se citar aquelas que utilizam: informações de bordas (WANG; OLIENSIS, 2010; CELIK; TJAHHADI, 2010); fragmentos (DALIRI; TORRE, 2009; LEVIN; WEISS, 2009); repetição (ZENG; GOOL, 2008); particionamento de árvores (WANG et al., 2008); modelo não-paramétrico Bayseano (ORBANZ; BUHMANN, 2008); modelo de contorno ativo geométrico sem re-inicialização (YING et al., 2009); campos aleatórios de Markov com crescimento de regiões (QIN; CLAUSI, 2010); campos aleatórios de Markov com corte em grafos (CHEN et al., 2010) e modelos locais Chan-Vese (LCV) (WANG; HUANG; XU, 2010). A maioria dos algoritmos lidam tanto com imagens coloridas quanto com imagens em tons-de-cinza. Alguns algoritmos de segmentação de imagens são especificamente destinados a imagens coloridas (CELIK; TJAHHADI, 2010; YU et al., 2010).

2.1.7 Agrupamento

O agrupamento de fotografias tem como propósito definir as associações entre grupos de fotografias. A associação pode ser definida por intermédio da informação semântica encontrada (o número de faces detectadas, a cor predominante, etc.), assim como de informações presentes nos metadados da imagem (EXIF), tal como a posição GPS da localidade onde a fotografia foi obtida.

Algoritmos para o agrupamento de fotografias são destinados a agrupá-las automaticamente por intermédio das características extraídas. Algoritmos fundamentados em grafos (FOGGIA et al., 2008), modelos discriminantes locais e integração global (LDMGI) (YANG et al., 2010) são utilizados para esta finalidade.

Os erros decorrentes do processo de agrupamento podem ser utilizados para realimentar tais processos de forma a melhorar os resultados. Para tanto, a interação com o usuário é utilizada, a fim de que se tenha uma resposta relevante sobre os erros do sistema (BRUNEAU; PICAROUGNE; GELGON, 2010).

Em se tratando de fotografias, a informação sobre um determinado evento, ao qual tais fotografias pertencem, pode ser relevante para a realização do agrupamento. A

partir da meta-informação presente no EXIF, podem-se obter informações de interesse. Obrador e Moroney (2009), propuseram um método para a organização hierárquica de fotografias fundamentada no horário em que a fotografia foi obtida. De acordo com a abordagem utilizada, ordenam-se as fotografias de forma crescente, por seu horário de aquisição e, em seguida, calcula-se a diferença entre pares de fotografias adjacentes, a partir da função d_t , explicitada na Equação 2.16.

$$d_t(t_i) = -\log((t_i - t_{i-1}) + 1), \quad (2.16)$$

em que t_i é a data/hora na qual a i -ésima fotografia foi obtida.

Uma vez calculadas todas as diferenças entre as fotografias adjacentes, gera-se uma árvore binária na qual o nó raiz corresponde à maior diferença existente naquele conjunto. Os nós seguintes são obtidos recursivamente, até um certo limiar pré-definido.

O horário da fotografia também foi utilizado por Wallick (2007). Para a identificação dos eventos, adotou-se o cálculo do triplo de tempo do espaçamento médio entre o tempo de fotografias subsequentes como heurística.

Na impossibilidade de utilização do horário da fotografia, Wu et al. (2011) utilizaram informações tais como a distância entre cores da fotografia, a partir do *Earth Mover's Distance* (EMD) (RUBNER; TOMASI; GUIBAS, 2000) e a distância euclidiana média entre as faces detectadas nas fotografias comparadas.

Por sua vez, Egorova e Safonov (2011) combinaram informações temporais e espaciais para realizar um agrupamento semi-automático das fotografias.

2.1.8 Discussão

Nesta seção, busca-se analisar, de forma crítica, as pesquisas destinadas à análise de fotografias ora discutidas. As pesquisas serão avaliadas quanto ao conjunto de imagens utilizado e o método de validação empregado.

Conjuntos de imagens

Na maior parte das pesquisas ora discutidas, o objetivo é a seleção automática de fotografias, similarmente ao modo humano de realizá-la. Assim, é fundamentalmente

importante garantir que a amostra das fotografias seja representativa nos testes.

Alguns estudos foram realizados para identificar o comportamento do usuário enquanto adquiriam (LUX; KOGLER; FABRO, 2010), compartilhavam (MILLER; EDWARDS, 2007), analisavam (FEDOROVSKAYA; NEUSTAEDTER; HAO, 2008), e gerenciavam fotografias (CUNNINGHAM; MASOODIAN, 2007). Entretanto, com base no estudo bibliográfico, não foi possível derivar evidências conclusivas sobre o modo de avaliação da maior parte dos algoritmos, uma vez que não há uma definição de procedimentos para a realização da análise subjetiva. Tal definição de procedimentos é essencial pois existem diversos fatores que podem influenciar a avaliação, assim como discutido anteriormente.

Pode-se concluir, com relação ao número de imagens e suas fontes, que não há um consenso na base de fotografias a serem utilizadas. Este fato impede uma comparação direta justa entre os resultados das pesquisas revisadas, além de impedir a reprodutibilidade dos experimentos. Também se pode concluir que nenhuma das pesquisas revisadas apresentou uma categorização da base de dados (e.g., não é possível saber a distribuição do número de participantes dos experimentos, no conjunto). Por fim, algumas das pesquisas revisadas, apresentaram apenas uma verificação visual dos resultados (e.g., Achanta e Süsstrunk (2009) e Banerjee e Evans (2004)), de modo que não há parâmetro de comparação com tais pesquisas.

Deve-se enfatizar a não utilização, na maior parte dos pesquisas, de uma base rotulada, pública e representativa para a análise de fotografias. Assim, muitos autores realizaram uma busca na *Web* por imagens que pudessem dar riqueza e diversidade ao conjunto (OLIVEIRA et al., 2002; SERRANO; SAVAKIS; LUO, 2002; CHEN et al., 2003; SHEN et al., 2009). Além disto, um problema fundamental reside na ausência de licenças para a utilização de imagens em experimentos. Existem bases que são disponíveis para pesquisas (e.g., Flickr (Flickr, 2013)) e outras que permitem a utilização a partir da licença *Creative Commons* (Creative Commons, 2012), apesar de ambas serem bases não rotuladas destinadas à avaliação de fotografias. Algumas bases foram utilizadas para a avaliação subjetiva (e.g., DPChallenge (Challenging Technologies, 2013) e Photo.net (GREENSPUN, 2013)). Entretanto, tais bases são destinadas a concursos de fotografias, não refletindo, portanto, a realidade da fotografia de consumo,

a qual possui menor qualidade e avaliadores menos exigentes.

Dois autores, dentre aqueles revisados, produziram bases com o intuito de disponibilizá-las para outras pesquisas. A primeira, criada por Luo, Wang e Tang (2011), contém 17.000 fotografias rotuladas. O conjunto foi construído com o objetivo de ser diverso, uma vez que as fotografias são sub-divididas em sete categorias e rotuladas como de alta ou baixa qualidade. O problema desta base de imagens reside no processo de rotulagem. Algumas informações não estão claras, tais como o número exato de votos por categoria e a origem e nível de conhecimento do fotógrafo e rotulante. Além disto, uma rotulagem mais precisa (ao invés de classificar apenas como alta/baixa qualidade) permitiria um uso mais genérico da base em problemas de análise e processamento de imagens. A segunda base, criada por Bhattacharya, Sukthankar e Shah (2010), apresenta um número reduzido de imagens (apenas 632). Outros fatores, tais como aqueles avaliados na abordagem de Luo, Wang e Tang (2011), não puderam ser avaliados, devido à impossibilidade de obtenção da base, em decorrência de um erro no servidor *Web* no qual está hospedada a base.

Assim sendo, não foi possível identificar estudos comparativos envolvendo diferentes abordagens que considerem bases de fotografias publicamente disponíveis. Este fato impossibilita uma comparação confiável e justa entre todas as pesquisas ora realizadas e estudadas sobre o tema discutido neste documento.

Validação

Esta seção contém uma discussão sobre métodos adotados na validação das abordagens revisadas. Uma vez que pode ser considerada uma forma de arte (HEDGECOE, 2009), não há uma forma simples e direta de decidir se uma fotografia é esteticamente agradável ou não. Assim, um aspecto importante a ser analisado é como abordagens afins àquela adotada nesta tese foram validadas.

Os métodos destinados para este fim podem ser classificados como subjetivos ou objetivos. Os métodos subjetivos envolvem usualmente a validação a partir de experimentos nos quais humanos são instruídos a opinar sobre critérios pré-definidos, relacionados às fotografias que estão sendo avaliadas. Diversas formas de avaliação (MANTIUK; TOMASZEWSKA; MANTIUK, 2012) podem ser adotadas, a saber:

- **Avaliação por estímulo simples.** O participante pontua uma fotografia ou conjunto de fotografias a partir de uma faixa de escore numéricos (e.g., 0-10) ou de semântica diferencial (e.g., excelente, boa, ruim e péssima). Durante a fase de avaliação, cada fotografia é tipicamente apresentada ao participante por um período fixo de tempo (e.g., 3 segundos);
- **Avaliação por estímulo duplo.** Análoga à avaliação por estímulo simples. Neste caso, a avaliação do participante se dá a partir de uma fotografia de referência.
- **Escolha forçada.** Ao invés de atribuir escore às fotografias avaliadas, apresenta-se ao participante uma sequência de pares de fotografias, em cada uma das quais deverá ser escolhida a fotografia julgada como sendo a de melhor qualidade do par apresentado; e
- **Julgamento pareado da similaridade.** Similar à escolha forçada, mas o participante deverá, além de escolher a fotografia que julgar de melhor qualidade, atribuir um escore à referida fotografia.

Outros detalhes e formas de comparação podem ser encontrados na pesquisa de Mantiuk, Tomaszewska e Mantiuk (2012), que realizaram uma comparação entre as quatro formas de julgamento supradescritas. Adicionalmente, os autores comentam que para a realização de comparações de algoritmos para a análise de fotografias, a escolha forçada se afigura a forma mais acurada e rápida de julgamento.

Outros fatores influenciam na avaliação realizada por humanos, a saber:

- **Número de participantes.** Uma vez que a opinião sobre a qualidade de uma fotografia pode variar entre participantes, é importante que haja uma amostra representativa de participantes, de modo a ser possível identificar padrões de comportamento significativos nos grupos observados;
- **Equipamento utilizado.** Durante a condução do experimento em um ambiente não-controlado, o equipamento utilizado no experimento pode interferir no resultado (e.g., calibração da tela quando se utilizam cores);

- **Conhecimento em fotografia.** Especialistas ou fotógrafos profissionais avaliam fotografias diferentemente dos consumidores ou fotógrafos amadores;
- **Diversidade cultural.** O estilo e alvo da fotografia podem influenciar no julgamento, dependendo do estilo e das origens do avaliador; e
- **Número de fotografias.** O número de fotografias no experimento é um fator tão crucial quanto o número de participantes. Se, por um lado, um grande número de fotografias pode representar bem a diversidade das fotografias, por outro lado, este pode reduzir o número de participantes voluntários no experimento, uma vez que o processo de julgamento será mais demorado.

Uma vez que, de acordo com a literatura revisada, não há uma base de dados que explicitamente considere tais fatores, pode-se alegar que a maior parte das conclusões derivadas de experimentos subjetivos apresentam vieses. Além disto, ainda que considerados tais fatores, não há um consenso quanto aos valores ideais. Portanto, grande parte das pesquisas revisadas apresentam decisões questionáveis quanto à etapa de validação, tais como o número de participantes (e.g., 3 participantes para a avaliação (OBRADOR, 2008)), conhecimento em fotografia (e.g., a maioria dos participantes são especialistas em fotografia (KE; TANG; JING, 2006)) e o número de imagens utilizadas (e.g., apenas 34 fotografias foram utilizadas para representar a população, na pesquisa de Fedorovskaya, Neustaedter e Hao (2008)).

Por outro lado, métricas objetivas de avaliação permitem uma avaliação mais padronizada. Por exemplo, os melhores algoritmos podem ser aqueles cujas taxas de falso positivo são as mais baixas, em um cenário de detecção de faces. Métricas objetivas são menos onerosas, dado que não dependem da disponibilidade e coerência dos participantes. Entretanto, existem características importantes que não são adequadamente avaliadas por algoritmos computacionais, tais como o aspecto visual global de uma fotografia. Mesmo humanos podem discordar de um resultado de classificação. Assim, ambas as abordagens (objetivas e subjetivas) são importantes, cada uma delas em um cenário específico de aplicação.

Apesar dos resultados relatados nas tabelas terem sido obtidos a partir de algoritmos e objetivos diferentes, é possível concluir que muitas abordagens optaram pela avaliação

subjetiva. Isto deve-se, provavelmente, à falta de consenso entre o conjunto de imagens a ser utilizado como *ground-truth* e da tarefa eminentemente subjetiva de comparar imagens.

Portanto, conclui-se que não há clareza com relação ao número de participantes considerado nos experimentos subjetivos, ao grau de familiaridade de tais participantes com o problema e à abordagem experimental adotada.

2.2 Diagramação automática de documentos

A diagramação de um documento consiste na disposição de um determinado conteúdo, constituído de elementos textuais, imagens, etc., em um dado espaço-alvo, e.g., uma página de tamanho A4, uma página *Web*. No processo de diagramação, além do posicionamento dos elementos, define-se também a fonte de texto utilizada, as cores dos componentes do documento e demais detalhes secundários a serem considerados no processo de diagramação do documento. Denomina-se *layout* (ou leiaute) o resultado de um processo de diagramação.

2.2.1 Fundamentação sobre diagramação

A seguir, são apresentadas definições para diversos termos, frequentemente adotados na literatura e referentes à diagramação de documentos, que serão utilizados ao longo desta pesquisa.

Quando o processo de distribuição de componentes visuais está associado a metas específicas, tais como a otimização do espaço e o direcionamento da atenção do observador, diz-se que o processo de diagramação é parte de um projeto gráfico (DRUCKER; MCVARISH, 2010).

Um projeto gráfico possui princípios básicos, os quais vêm sendo adotados desde a Grécia antiga, a exemplo da proporção áurea, adotada por Fídias no projeto do Partenon, nos cânones de beleza artística (arquitetura e pintura) gregos e renascentistas e, mais recentemente, em processos de diagramação os mais diversos. Os princípios em questão levam em consideração aspectos da psicologia humana, tendo alguns deles sido objetos de estudos e experimentos que reforçam a eficácia de sua adoção. Entretanto,

não se pode tornar tais regras obrigatórias, uma vez que é possível subvertê-las em algumas circunstâncias sem o comprometimento dos resultados obtidos. Portanto, os princípios básicos de um projeto gráfico devem ser tão somente encarados como guias e sugestões para se atingir um determinado objetivo (DRUCKER; MCVARISH, 2010).

A partir da utilização dos princípios básicos de um projeto gráfico, é possível que o resultado seja considerado mais atrativo. Neste documento, considera-se um projeto gráfico como atrativo quando este desperta a sensação de prazer estético e conforto visual em um dado universo de observadores. Em contrapartida, será considerado não-atrativo aquele que seja desconfortável ao olhar e cuja visualização não seja associada à sensação de prazer estético, não interessando, *a priori*, o motivo.

A atratividade, neste documento, também empregada como sinônimo de agradabilidade, não é a única métrica adotada no processo de avaliação de um projeto gráfico. Por exemplo, um projeto gráfico pode ser avaliado por sua capacidade de direcionar a atenção de um dado universo de observadores para um ponto de interesse, e.g., uma propaganda ou uma logomarca (DRUCKER; MCVARISH, 2010).

Para a realização de um projeto gráfico, é essencial que sejam seguidos princípios básicos. Tais princípios servem para orientar os profissionais da área na obtenção de resultados consistentes e visualmente atrativos. Portanto, seguir estritamente as regras não é garantia de um projeto gráfico de alta qualidade. A não conformidade aos princípios não implica projetos gráficos inferiores. Apesar disto, a observação dos princípios básicos é um bom ponto de partida para um projeto gráfico de alta qualidade. Contudo, deve-se evitar a obediência excessiva aos princípios sob o risco de se produzir um projeto gráfico artificial e repetitivo.

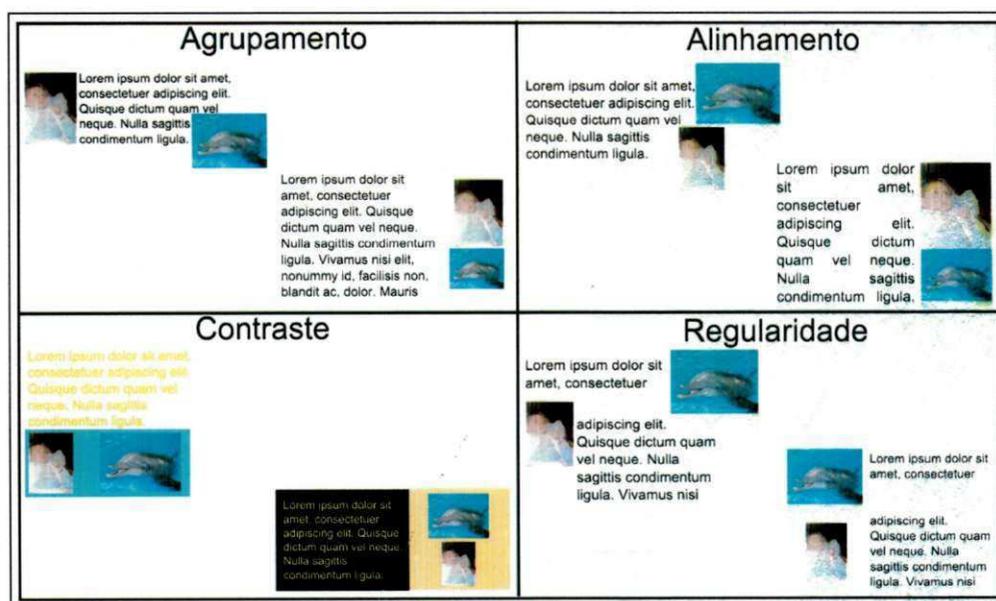
Dentre os princípios básicos de um projeto gráfico, destacam-se:

- **Agrupamento.** A proximidade de dois ou mais elementos de uma composição gráfica tende a determinar a unidade entre eles. O propósito básico da proximidade, de acordo com Williams (2003), é a organização. O simples agrupamento de elementos cria essa organização. De acordo com Gomes Filho (2000), “Proximidade e semelhança são dois fatores que, muitas vezes, agem em comum e se reforçam mutuamente, tanto para constituírem unidades como para unificarem a forma”.

- **Contraste.** “O contraste é o efeito que permite ressaltar o peso visual de um ou mais elementos ou zonas de uma composição mediante a oposição ou diferença apreciável entre elas, permitindo atrair a atenção do espectador para eles” (MORENO, 2007a).
- **Alinhamento.** Os objetos de um documento devem ser arranjados em uma página de tal forma que certo número de suas bordas estejam alinhadas, a fim de oferecer conectividade visual. Como recurso auxiliar ao alinhamento, podem-se citar as linhas de alinhamento, as quais determinam a distribuição alinhada dos objetos em relação ao resto do documento, podendo ser horizontais e/ou verticais (MORENO, 2007b).
- **Regularidade.** “A regularidade constitui o favorecimento da uniformidade dos elementos, e o desenvolvimento de uma ordem inspirada em algum princípio ou método constante e invariável” (DONDIS, 1991).

Na Figura 2.6, são ilustrados os princípios básicos supradescritos.

Figura 2.6: Ilustração dos princípios básicos de um projeto gráfico.



Como o objeto de estudo a ser diagramado nesta pesquisa é um álbum fotográfico, que consiste em um documento, alguns conceitos básicos sobre diagramação de docu-

mentos são descritos a seguir. Com relação à disposição da informação, os documentos podem ser classificados em contínuos e não-contínuos. Documentos contínuos são aqueles que seguem uma ordenação lógica, apresentando uma sequência ou percurso a ser seguido para se compreender o conteúdo do documento, e.g., livros, jornais, textos acadêmicos em geral. Por sua vez, documentos não-contínuos não apresentam nenhum requisito quanto ao posicionamento das informações, e.g., folhetos publicitários e alguns documentos bancários (BALINKSY; HOWES; WILEY, 2009).

Em geral, nos documentos são dispostos tanto conteúdos estáticos quanto dinâmicos. Os conteúdos estáticos são aqueles que ocupam posições fixas na página (e.g., cabeçalho, rodapé, logotipos, títulos). Por sua vez, os conteúdos dinâmicos são aqueles cujas posições podem ser alteradas, a fim de conferir diferentes aspectos visuais à página na qual se inserem.

Com relação ao conteúdo dinâmico, sua área é virtualmente dividida em espaços nos quais são posicionadas as informações. Estes espaços usualmente possuem forma retangular e são denominados contêineres (BALINKSY; HOWES; WILEY, 2009). Quando da utilização de conteúdos estáticos na diagramação de um documento, estes possuem contêineres pré-definidos, sendo os contêineres restantes destinados aos conteúdos dinâmicos. Os contêineres também podem ser côncavos, possuindo regiões curvadas ou não retangulares, assim como podem apresentar áreas de descontinuidade, i.e., áreas vazias entre elementos, o que dificulta o processo de diagramação.

Um exemplo para ilustrar melhor estes conceitos é um documento que contém figuras, textos e propagandas. O documento pode ser dividido em regiões estáticas e dinâmicas. Nas regiões estáticas, estão posicionados o título, o número da página e as bordas dentre outros possíveis elementos constituintes. Na região dinâmica, serão dispostos os demais conteúdos. Uma certa parte da região dinâmica poderá ser destinada a propagandas. Sendo assim, essa área terá contêineres estáticos para propagandas. Em outra região, poderão existir contêineres estáticos para o texto. Por fim, poderão existir contêineres dinâmicos para as figuras e textos auxiliares, os quais serão moldados de acordo com os demais conteúdos da página.

A diagramação automática de documentos afigura-se um grande desafio, uma vez que muitos fatores precisam ser levados em consideração, além do posicionamento do

conteúdo, dentre os quais podem-se destacar: a interligação entre as informações, o formato da informação e o impacto esperado com a apresentação da informação.

2.2.2 Pesquisas sobre diagramação automática de documentos

A pesquisa sobre diagramação automática de documentos documentada na literatura apresenta abordagens destinadas à diagramação de diferentes tipos de conteúdo (e.g., fotografias e texto) em mídias de diferentes tipos e tamanhos (e.g., páginas *Web*).

Kröner (1999) propôs o *DesignComposer*, uma ferramenta destinada à diagramação automática de páginas *Web*. O *DesignComposer* resultou de uma combinação de diferentes técnicas de inteligência artificial (IA) que utilizam um modelo de conhecimento declarativo gerado por especialistas ou pelo próprio usuário.

A entrada principal da ferramenta supracitada é um modelo de apresentação que utiliza uma linguagem de marcação (usualmente XML ou HTML), podendo ser gerado automaticamente por uma aplicação de *software* externa ou manualmente por um usuário. Outro dado de entrada utilizado é a especificação de um *layout*, que é armazenado em uma base de conhecimento.

As tarefas de diagramação são realizadas em três etapas: primeiro, ocorre a preparação da entrada a ser processada e da base de conhecimento em *layouts*. Em seguida, a criação do *layout* é realizada pelo módulo gerador. Finalmente, o resultado é pós-processado por meio da interação com o usuário. De acordo com Kröner (1999), o modelo do objeto gerado, em virtude das diferentes técnicas de inteligência artificial utilizadas, permite grande flexibilidade e facilidade de manutenção. Não há nenhuma indicação de que o resultado do sistema tenha sido sistematicamente verificado ou testado.

Similarmente à abordagem anterior, Purvis (2002) propôs a utilização de algoritmos genéticos (AG) (LINDEN, 2008) para a diagramação automática de documentos, não se limitando, porém, à produção de páginas *Web*. Segundo a autora, as abordagens para a criação de documentos personalizados existentes à época da publicação ou não eram automáticas ou se restringiam a um tipo específico de documento. Deste modo, a principal vantagem de tal abordagem, conforme a autora, reside na possibilidade de se realizar a diagramação em qualquer tipo de documento alvo, sem a necessidade de

grandes modificações na estrutura básica do sistema.

Na solução original proposta por Purvis (2002), a diagramação de documentos era modelada a partir de restrições simples, e.g., alinhamento do texto e largura dos conteúdos. Todavia, de acordo com a autora, os resultados obtidos não apresentaram um aspecto visual agradável, além de requererem especificações de restrições em linguagem de baixo nível, onerando o processo de escolha dos valores limítrofes para tais restrições.

Para amenizar este problema, AG passaram a ser utilizados, uma vez que possibilitariam a combinação sistemática de valores para as restrições. De acordo com a autora, o uso de AG permitiu a obtenção de resultados mais atrativos, restrições mais simples e uma abordagem mais flexível com relação ao tipo e dimensões do documento alvo. De modo similar a abordagem de Purvis (2002), outras abordagens foram propostas para a diagramação de documentos a partir da otimização realizada pelos AG (GEIGEL; LOUI; LOUI, 2001; GEIGEL; LOUI, 2003)

Seguindo a linha proposta por Purvis (2002), Balinsky, Howes e Wiley (2009) propuseram um algoritmo destinado à diagramação automática de documentos não contínuos. Tal algoritmo se fundamenta em um modelo generalizado, o qual contém uma mistura de objetos estáticos e espaços, denominados pelos autores *magic spots*. Os *magic spots* servem para reservar espaço a conteúdos tais como imagens-chave ou últimas ofertas. O espaço restante é particionado em um conjunto de contêineres. O conteúdo dinâmico é disposto em regiões da página que podem sofrer expansão ou contração, de acordo com as necessidades do conteúdo. Tais regiões são classificadas como itens expansíveis e são armazenados em contêineres. A abordagem descrita pelos autores compõem-se de oito etapas: pré-processamento dos dados, construção dos conteúdos, determinação dos conteúdos estáticos e *magic spots*, divisão da página em contêineres, seleção dos contêineres, agrupamento dos contêineres, busca de soluções simultâneas para cada contêiner e seleção da solução a ser usada.

Uma limitação evidente da abordagem supradescrita é o fato de serem utilizados apenas espaços retangulares, o que reduz sua flexibilidade. A fim de permitir a diagramação de um universo mais abrangente de documentos, esta abordagem poderia ser estendida, de modo a incluir a distribuição de objetos em regiões não retangulares (e.g., regiões circulares).

Em uma outra abordagem similar àquela adotada por Purvis (2002), Sridharan, Sundaram e Brungart (2004) descreveram um mecanismo a partir do qual livros de exercícios (*workbooks*) podiam ser gerados automaticamente, a partir dos objetivos e preferências dos usuários, assim como de princípios de projeto gráfico.

Os autores descreveram um estudo de caso envolvendo o uso do sistema em um ambiente eletrônico, com o propósito de auxiliar crianças no processo de aprendizagem de conceitos de geografia. Uma vez que o usuário interage com o mundo eletrônico, seu perfil é continuamente atualizado em função de suas interações com o sistema. Assim, novos conhecimentos podem ser inseridos, bem como mantidos, em função do processo interativo usuário-sistema. Foram incorporados à diagramação automática elementos de projeto gráfico, tais como contraste e proximidade.

A abordagem descrita por Sridharan, Sundaram e Brungart (2004) possui um diferencial em relação às abordagens anteriores, devido à utilização de um processo interativo envolvendo elementos virtuais (programas de computador) e reais (livros, folhas de papel). Esta interação é muito importante para um conceito educacional, pois prevê uma interação entre o mundo real e o mundo virtual, resultando em um ciclo meio eletrônico-papel.

Por sua vez, Lok e Kan (2003) propuseram uma abordagem para melhorar a exibição dos resultados de um sistema de recuperação de informação (*information retrieval - IR*). Para tanto, algoritmos de diagramação automáticos foram usados como etapa de pós-processamento com fins à recuperação da informação, dando origem ao *Centrifuser*.

O *Centrifuser* é um sistema de sumarização de informação visual que possui o intuito de satisfazer as necessidades dos navegadores e dos pesquisadores. O *Centrifuser* adota a seguinte sequência de etapas: primeiro, os documentos são convertidos em árvores de tópicos; em seguida, cálculos de similaridade entre árvores estruturadas são utilizados para encontrar os tópicos similares e distintos; por fim, o módulo de sumarização recebe os tópicos similares e realiza a extração das sentenças marcantes para a sinopse do documento. O sistema proposto pelos autores permite a renderização da informação em dispositivos de diferentes resoluções, sendo possível, por exemplo, a visualização de um mesmo conteúdo em um dispositivo móvel (e.g., PDA) e em um computador do tipo *desktop*, sem que haja interferência do usuário. A solução para o

sistema de recuperação de informação adota um sistema de navegação olho de peixe (*fish-eye*), no qual a adaptação do texto à tela de um dispositivo se dá pela variação no comprimento real do texto, ao contrário de outros sistemas, nos quais são realizadas modificações em parâmetros tais como o tamanho da fonte.

Diferentemente das pesquisas supracitadas, Xiao et al. (2008) propuseram uma abordagem para a criação de fotomontagens: o sistema *miCollage*, a partir do qual é possível realizar fotomontagens automaticamente ou com base em alterações definidas pelo usuário. Esta possibilidade de interação, após a geração da fotomontagem, representa uma diferenciação importante no tocante às abordagens anteriores, pois pode reduzir as frustrações do usuário em circunstâncias nas quais o processamento automático não produza os resultados desejados. O sistema consiste de três módulos principais: seleção, edição e diagramação. No módulo de seleção, o sistema cria sugestões de quais fotografias devem ser submetidas ao processo de fotomontagem e também auxilia o usuário a identificar fotos similares ou relacionadas. No módulo de edição, o sistema aplica efeitos de corte e realce automático das tonalidades e cores presentes nas fotografias pré-selecionadas. Por fim, no módulo de diagramação, são providas sugestões de *layouts* alternativos, bem como são aceitas alterações dos usuários para criação de *layouts* personalizados.

Atkins (2008) classificou as restrições para a diagramação de documentos em *primárias* e *secundárias*, definindo como *primárias* aquelas que precisam ser respeitadas, e como *secundárias* aquelas que se deseja que sejam respeitadas. Portanto, as restrições primárias definidas pelo autor foram: (1) a relação de aspecto das fotografias deveria ser preservada e (2) as bordas e espaçamento entre fotografias adjacentes deveriam ser especificadas. Também foram duas as restrições secundárias definidas pelo autor: (1) as áreas das fotografias deveriam ser proporcionais às áreas relativas previamente informadas e (2) as fotografias deveriam ocupar o máximo de área possível do espaço disponível.

A abordagem descrita por Atkins (2008), denominada BRIC (*Blocked Recursive Image Composition*), fundamenta-se na criação de árvores binárias, nas quais cada nó consiste em uma partição horizontal ou vertical do espaço disponível e cada folha consiste em uma imagem que ocupará o espaço de tais partições. Caso k fotos sejam

utilizadas, o algoritmo necessita de k iterações, sendo que em cada uma incrementa-se o número de imagens a serem diagramadas. Durante cada iteração, diversas árvores são criadas, variando-se cada partição em horizontal e vertical. Para a iteração seguinte, escolhem-se as L melhores árvores a partir de uma função de avaliação da árvore. Este processo se dá pela solução de um sistema linear de equações no qual calcula-se as dimensões que cada imagem precisa ter de modo que o conjunto de fotografias ocupe o máximo de área disponível.

No *miCollage*, o foco da aplicação é um algoritmo que apresente precisão na identificação da similaridade entre as fotografias selecionadas e velocidade na criação das fotomontagens. Com o *miCollage*, dois cenários de seleção de fotos são atendidos. No primeiro cenário, o “povoamento automático” do documento (*auto-populate*), a meta é a geração automática de uma diagramação completa. O segundo cenário, “incremental”, visa à seleção de novas fotografias em um grupo de fotos que ainda não tenha sido utilizado no processo de diagramação e sua inclusão no documento. No processo de avaliação descrito pelos autores, foram avaliados, a partir de um questionário aplicado aos usuários do sistema, o tempo de criação de um álbum por parte do usuário, o índice de satisfação com a escolha das fotografias e com o resultado final.

Piccoli (2010), em uma abordagem similar àquela de Balinsky, Howes e Wiley (2009), propôs um método destinado à diagramação automática de documentos de conteúdo variável. Os documentos criados seguem um *layout* hierárquico, dividido em regiões retangulares e alinhadas em colunas. A abordagem de Piccoli (2010) diferencia-se das demais revisadas por não utilizar métodos aproximativos (e.g., algoritmos genéticos) para associar regiões de uma página aos elementos de um documento. A escolha de um algoritmo determinístico foi feita devido às necessidades de desempenho e confiabilidade exigidas no contexto de impressão de dados variáveis (*variable data printing* - VDP), que foi uma das aplicações da pesquisa. O termo VDP é usado para designar o processo automático de criação de documentos com conteúdo personalizado. Esta abordagem se fundamenta na realização de divisões sucessivas em uma página e da alocação de regiões de acordo com a área exigida para cada elemento (conteúdo), ao invés de utilizar geometrias fixas para os objetos, como ocorre em outras abordagens na área de diagramação.

No algoritmo desenvolvido por Piccoli (2010), a escolha da divisão da página obedece a critérios estéticos importantes para o projeto de documentos, tais como alinhamento e regularidade, bem como a ordem de leitura definida pelos elementos e a possibilidade de agrupamento dos elementos inter-relacionados. Além disto, o algoritmo desenvolvido realiza a distribuição do conteúdo em documentos compostos por várias páginas.

Em uma outra abordagem, Sandhaus, Rabbath e Boll (2011) propuseram um método para a diagramação fundamentado em princípios estéticos e destinado à diagramação de fotografias e textos. Sob determinadas condições, também é possível realizar o recorte e a diagramação das fotografias, partindo-se de uma estrutura de árvore. Entretanto, em algumas circunstâncias, utilizam-se leiautes pré-definidos, limitando em parte a abordagem.

Por sua vez, Rabbath, Sandhaus e Boll (2011, 2010) apresentaram abordagens fortemente relacionadas a esta pesquisa. Tais abordagens integraram conhecimento de diagramação e princípios de projeto gráfico para a produção de álbuns fotográficos. Foram definidas regras para agrupamento e seleção. O agrupamento é feito com base em anotações dos usuários e informações obtidas tanto da câmera (como datas e horas) quanto do aplicativo de uma rede social (e.g., número de comentários da foto e indivíduos que mais interagem com o usuário). Quanto ao processo de diagramação, este se organiza em distribuição do conteúdo (separação das fotografias em páginas), seleção do plano-de-fundo e obtenção da diagramação em alto nível, a qual é realizada por intermédio de restrições otimizadas por intermédio de AG.

Kuhna, Kivelä e Oittinen (2012) propuseram uma abordagem para a diagramação automática de revistas com respeito ao conteúdo. São utilizadas regras de avaliação de fotografias para a seleção e posicionamento das fotografias. Também se utilizaram as cores das fotografias componentes para a decisão de colorização dos destaques de texto do documento.

Bergmann et al. (2013) também afirmam que a produção automática de álbuns fotográficos pode ser vantajosa para usuários que desejem uma criação rápida de tais produtos. Entretanto a distribuição das fotografias ao longo das páginas do álbum não é feita automaticamente, forçando, portanto, a necessidade de intervenção do usuário.

Por sua vez, Egorova e Safonov (2013) propuseram algoritmos adaptativos para a produção de fotolivros. Os autores também adotaram uma abordagem de produção de álbuns fotográficos fundamentada na sequência: seleção, agrupamento e diagramação. Até o momento, tal abordagem consiste em um grupo de algoritmos isolados, não havendo integração entre eles.

Finalmente, Xue et al. (2013) propuseram uma abordagem que também utiliza o BRIC (ATKINS, 2008) como algoritmo para o leiaute de álbuns fotográficos. O sistema desenvolvido pelos autores permite identificar o relacionamento entre as fotografias a partir de uma abordagem de seleção iterativa em que as fotografias são agrupadas a partir do algoritmo *k-means*.

2.2.3 Discussão

Analisando-se as pesquisas relacionadas, pôde-se perceber lacunas quanto ao formato do leiaute, dado que parte de tais pesquisas requerem a exigência de um leiaute pré-definido (e.g., na abordagem proposta por Sandhaus, Rabbath e Boll (2011)). Para outras pesquisas, permite-se alterações nos leiautes existentes (e.g., na abordagem proposta por Xiao et al. (2008)). Em ambos os casos, há uma limitação quanto ao resultado final, uma vez que não há diferenças significativas nos leiautes produzidos.

Por outro lado, também foram propostas pesquisas em que se dá mais liberdade ao leiaute produzido, na medida em que testam-se diversas opções de leiaute, as quais são avaliadas por uma função custo a qual é parametrizada por restrições previamente definidas, sendo escolhido o leiaute com menor custo para o conjunto de restrições. Neste tipo de abordagem, popularizou-se a utilização de AG (e.g., as pesquisas de Purvis (2002), Geigel e Loui (2003) e Geigel, Loui e Loui (2001)). Em virtude do formato de evolução utilizado, as abordagens pautadas em AG tornaram-se lentas. Para evitar as desvantagens causadas pelos AG, foram empregados, no processo de leiaute, métodos não aproximativos (e.g., a pesquisa realizada por Piccoli (2010)).

Assim, pode-se concluir que, de forma geral, a grande deficiência das abordagens supra-mencionadas é a existência de um formato pré-concebido ou o tempo necessário para o processo de otimização.

2.3 Recorte automático de fotografias

Nesta pesquisa, são considerados como algoritmos para o redimensionamento automático de imagens aqueles destinados à remoção de *pixels* de uma fotografia, de modo a adaptá-la a uma dimensão-alvo pré-determinada. Apesar de não ser um requisito, espera-se que as dimensões finais da fotografia sejam equivalentes à maior região retangular que esteja em conformidade com a proporção-alvo. Tal objetivo serve para maximizar a possibilidade de preservação do conteúdo e composição originais da fotografia.

Abordagens vêm sendo propostas na literatura com o objetivo de redimensionar imagens com base em seu conteúdo (*content-aware image resizing*). Podem-se definir duas categorias de abordagens para tal redimensionamento: (1) o recorte (*cropping*) e (2) o redirecionamento (*retargeting*). Outras taxonomias também são adotadas nas abordagens de redimensionamento com base no conteúdo propostas na literatura da área (e.g., Vaquero et al. (2010)). Entretanto, a classificação adotada nesta tese relaciona-se à manutenção dos elementos constituintes da fotografia. Nas abordagens fundamentadas no recorte de imagens, linhas e/ou colunas de *pixels* da imagem são removidas, até que a proporção-alvo seja alcançada. Por sua vez, nas abordagens fundamentadas no redirecionamento, os *pixels* a serem removidos não precisam, necessariamente, pertencer a uma mesma linha ou coluna nem estarem em uma posição específica (e.g., próximos às bordas). Para que o recorte ocorra são escolhidos *pixels* conectados de uma extremidade à outra da imagem, sendo a escolha realizada por intermédio de uma função de minimização (ou maximização) de energia.

Portanto, enquanto a primeira categoria de abordagem (recorte) permite evitar a degradação do conteúdo da fotografia, a segunda (redirecionamento) possibilita um redimensionamento menos restritivo. Por conseguinte, as abordagens de recorte são especialmente úteis quando deseja-se preservar tanto a proporção original quanto o conteúdo da fotografia, enquanto as abordagens de redirecionamento são indicadas quando pequenas perdas e distorções são irrelevantes ou imperceptíveis. Vale enfatizar que, para uma mesma dimensão-alvo, ambas as abordagens removerão a mesma quantidade de *pixels*, alterando-se tão somente as coordenadas dos *pixels* que serão

removidos.

Uma revisão do estado da arte das categorias de abordagens fundamentadas no recorte e no redirecionamento é apresentada, respectivamente, nas duas próximas subseções. Em seguida, é apresentada uma discussão sobre a adequação de cada técnica ao problema discutido nesta pesquisa.

2.3.1 Algoritmos de recorte

Suh et al. (2003) propuseram um método para o recorte de miniaturas que funciona diferentemente para fotografias com e sem alvos humanos, no qual um detector de faces é executado. Uma vez que faces tenham sido detectadas, estas são recortadas com o auxílio de um retângulo que engloba todas as faces, mas que apresenta espaço das faces até as bordas da imagem. Se nenhuma face for encontrada, é utilizada a saliência para identificar quais são as regiões mais importantes da imagem. Os autores afirmaram que as fotografias são mais fáceis de serem identificadas em uma ampla base de imagens a partir da miniatura gerada pelo método proposto. A combinação entre faces e saliência não foi testada e os autores usaram três conjuntos de imagem: um conjunto com animais; um conjunto contendo faces; e a base de imagens Corbis (CORBIS, 2001-2009). Os resultados indicaram que, em fotografias contendo alvos humanos, o método utilizando o detector de faces permitiu uma identificação mais precisa do conteúdo da fotografia.

Santella et al. (2006) propuseram uma abordagem para o recorte semiautomático de fotografias. De acordo com os autores, a direção do olhar de um observador pode ser capturada por meio de uma câmera especial, de modo a auxiliar no recorte da fotografia. As áreas nas quais o usuário tiver focado o olhar por maior tempo, são preservadas. Em uma avaliação subjetiva, o método se mostrou tão eficiente quanto o recorte manual. Apesar dos resultados serem muito bons, o método não é completamente automático, requerendo (1) a intervenção do usuário no recorte de cada fotografia e (2) um equipamento especial para a captura da direção do olhar do observador.

Ciocca et al. (2007) previram a identificação da classe a qual pertence a fotografia (i.e., *close-up*, paisagem ou outra) para efetuarem automaticamente o recorte da região com base na classe identificada. Os autores mencionaram uma taxa de 53% de aprova-

ção por parte dos usuários do sistema com relação ao recorte. Entretanto, os autores também reportaram problemas no algoritmo detector de faces que é utilizado na etapa de identificação da classe da fotografia.

2.3.2 Algoritmos para redirecionamento de imagens

Chen et al. (2003) propuseram um método para selecionar regiões importantes em uma fotografia de modo a exibir tais regiões em telas de pequenas dimensões e.g., celulares e dispositivos móveis. Os autores utilizaram detectores de face, detectores de texto e atenção visual, todos combinados. No método proposto, objetos de atenção (*attention objects* - AOs) foram encontrados, com o intuito de identificar elementos perceptíveis. Cada AO é composto por regiões de interesse (*regions of interest* - ROI), um valor de atenção (*attention value* - AV) e um valor mínimo para percepção (*minimal perceptible size* - MPS). O valor de atenção é calculado a partir da Equação 2.17, em que AV_i^k denota a atenção do objeto i no modelo k e w_k é o peso do modelo.

$$AV_i = w_k \times \frac{AV_i^k}{\sum_i AV_i^k} \quad (2.17)$$

Entretanto, os autores não descreveram como os pesos para cada AV foram obtidos. O método foi avaliado a partir de um experimento subjetivo no qual foram utilizadas 56 imagens e que teve participação de 15 observadores humanos. Segundo os autores, os resultados refletiram a aceitação por 71% dos participantes da avaliação.

A pesquisa de Setlur et al. (2005) é fortemente relacionada aquela desenvolvida por Chen et al. (2003). Setlur et al. (2005) propuseram um modelo no qual se utiliza a atenção visual e a detecção de faces, as quais são combinadas a partir de uma métrica denominada valor de importância (IV), descrita pela Equação 2.18.

$$IV_i = w_k \times \overline{IV_i^k}, \quad (2.18)$$

em que $\overline{IV_i^k}$ é o valor de importância normalizado de IO_i detectado no modelo k e w_k é o peso associado.

Da mesma forma que na abordagem proposta por Chen et al. (2003), não foi descrito

em detalhes como os pesos foram obtidos, apesar de ser descrito que a otimização dos pesos foi realizada a partir da resposta de usuários do sistema.

Avidan e Shamir (2007) propuseram uma abordagem para redimensionar imagens com respeito ao conteúdo, denominada *seam-carving*, na qual os *pixels* são removidos de uma imagem a partir de uma função de energia que permite calcular quais regiões possuem pouca importância quanto ao conteúdo. Tal abordagem foi posteriormente agregada a um *software* comercial de referência na área de edição de imagens (Adobe Photoshop CS5) com o nome de “content aware scaling” (INC., 2010).

Achanta e Süsstrunk (2009) desenvolveram um método para o cálculo da saliência, diferentemente do método tradicional de Itti, Rees e Tsotsos (2005), de modo que a saliência pudesse ser utilizada como função de energia para o método de *seam-carving*.

O outro método que utiliza *seam-carving*, mas preserva a composição da fotografia, foi proposto por Liu et al. (2010). Tal método combina o uso de saliência (ITTI; REES; TSOTSOS, 2005), linhas e regras de composição fotográfica para a decisão da área a ser recortada.

Hua, Li e Zhong (2011) propuseram um algoritmo para cálculo da similaridade entre regiões da imagem utilizando uma medida de similaridade bidirecional, de modo a julgar as modificações realizadas pelo *seam-carving* evitando, assim, grandes distorções.

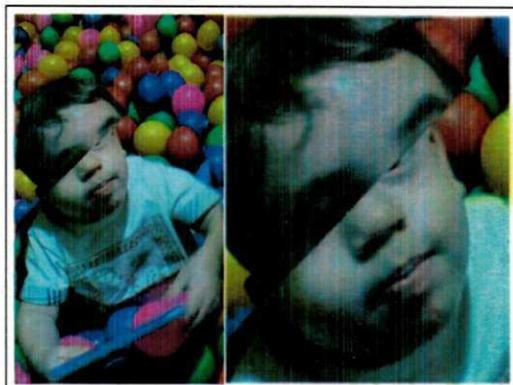
Apesar dos resultados promissores que podem ser obtidos pela abordagem *seam-carving*, existem casos nos quais fortes distorções são produzidas mesmo havendo interferência humana.

O efeito ora descrito frequentemente acontece quando o alvo ocupa uma larga região na imagem. A distorção do alvo ocorrerá quando não existirem *pixels* suficientes para o recorte em regiões não ocupadas pelo alvo da fotografia. Pequenas distorções também são percebidas na ocorrência de linhas retas na imagem.

Naturalmente, em algumas situações específicas tais como quando o alvo não ocupa uma larga área da fotografia, o processo de redirecionamento da imagem pode ser utilizado, mas sem garantias de preservação do alvo a menos que exista interferência humana.

Na Figura 2.7, é ilustrado um exemplo do efeito da distorção que pode ser causado pela maioria dos métodos de *seam-carving*.

Figura 2.7: Exemplo de distorção causada por métodos de redirecionamento.



2.3.3 Discussão

Ao analisar as abordagens relacionadas, conclui-se que existe uma lacuna quanto aos métodos destinados ao recorte de fotografias contendo alvos humanos, mas que levem em consideração aspectos fotográficos, e.g., a composição fotográfica.

Tais métodos, apesar de serem úteis, por exemplo, para a exibição de fotografias em pequenas telas e sumarização de seu conteúdo, podem não ser apropriados caso o objetivo seja a impressão das fotografias, uma vez que a distorção é facilmente percebida em fotografias impressas (mais largas do que uma tela e normalmente visualizadas por um período maior de tempo).

Assim, para fins de produção de mídia impressa, se afiguram mais adequadas soluções cujos resultados sejam isentos de distorções. Nesta pesquisa, portanto, é proposta uma abordagem para o recorte de fotografias que não apresenta distorções em seu resultado. Três métodos foram escolhidos para comparação com a abordagem proposta: (1) o método de Suh et al. (2003), categorizado como um algoritmo de recorte; (2) o método de Avidan e Shamir (2007), como abordagem para *retargeting*; e (3) o método *Automatic Zoom & Crop* (AZ&C).

O método AZ&C recorta a imagem na área definida pelo maior retângulo que circunscribe uma região. Este retângulo deve estar centralizado na imagem. Tal método

pode parecer simples demais para ser capaz de produzir bons resultados. De fato, não se trata de um método sofisticado. Contudo, vem sendo amplamente utilizado pela indústria de revelação e impressão de fotografias.

Os recentes avanços nas tecnologias das câmeras fotográficas e a facilidade de obtenção de material instrucional sobre fotografia permitem o surgimento de um novo cenário, no qual as fotografias obtidas são mais bem compostas. Assim, a grande motivação para o desenvolvimento de um método de recorte para as fotografias, é a possibilidade de ajustá-las para uma nova proporção-alvo. Desta forma, pretende-se evitar o comprometimento da composição fotográfica originalmente proposta.

2.4 Considerações Finais

Neste capítulo, foi apresentada uma revisão bibliográfica acompanhada de fundamentação para as questões tratadas nos capítulos seguintes desta pesquisa.

A partir das pesquisas estudadas, pôde-se perceber que ainda existem diversas lacunas tanto na análise de fotografias digitais quanto na diagramação de álbuns fotográficos.

Com relação à análise de fotografias digitais, percebe-se que poucos são as abordagens que avaliam fotografias, de forma objetiva ou subjetiva, com vistas a estabelecer uma correlação da avaliação automática com a avaliação humana.

Capítulo 3

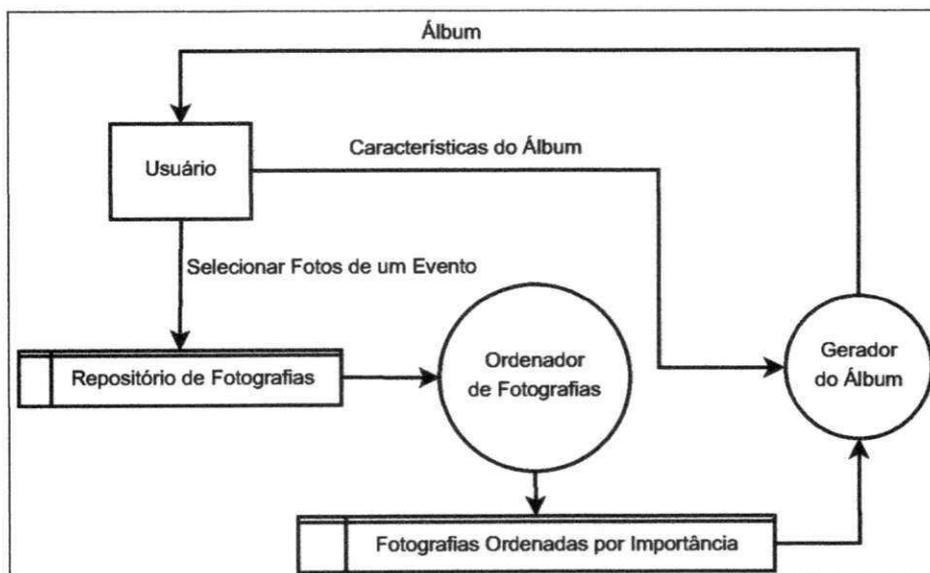
Sistema proposto

Neste capítulo, é descrito o sistema proposto nesta tese de doutorado. Este capítulo está estruturado da seguinte forma: na Seção 3.1, é apresentada uma visão de alto nível do sistema proposto. Na Seção 3.2, é descrita formalmente a solução proposta. Por fim, na Seção 3.3 são apresentadas as considerações finais.

3.1 Visão geral do projeto

De forma geral, o sistema proposto pode ser ilustrado conforme o diagrama de fluxo de dados (DFD) da Figura 3.1.

Figura 3.1: Diagrama geral do sistema proposto.



No diagrama, a entidade *usuário* é responsável pela criação de um *repositório de fotografias*, o qual deve ser composto por fotografias associadas a um único evento. Por evento, entende-se um acontecimento (e.g., temporal ou individual) o qual motivou a obtenção das fotografias. A criação de tal repositório não é parte do sistema proposto, sendo responsabilidade do usuário.

O sistema proposto, representado na Figura 3.1, pode ser dividido em três etapas fundamentais: seleção, agrupamento e diagramação das fotografias.

A primeira etapa, seleção, tem como primeiro passo a ordenação do conjunto de fotografias, de acordo com regras estipuladas, a exemplo daquelas introduzidas na pesquisa de Cavalcanti (2007). Assim, o conteúdo do repositório de fotografias será utilizado como entrada para o módulo *ordenador de fotografias*, o qual por sua vez, utiliza um conjunto de módulos avaliadores de fotografias como base para a ordenação. Os módulos avaliadores têm como objetivo analisar, de acordo com algum aspecto, uma fotografia e atribuir-lhe um escore que quantifique a sua qualidade. Não há garantia, contudo, de que um avaliador seja capaz de reproduzir a opinião do usuário quanto à qualidade global da fotografia. De modo a minimizar este problema, os módulos avaliadores foram combinados. A tarefa de análise e a combinação dos módulos avaliadores compõem o módulo *ordenador de fotografias*.

A segunda etapa, agrupamento, inicia-se a partir dos resultados da etapa de seleção. Utiliza-se dos metadados de data e hora em que as fotografias foram obtidas, com o intuito de identificar os subeventos associados às fotografias. Combinam-se, então, os resultados do agrupamento e da seleção de modo que, para cada subevento, apenas as melhores fotografias sejam utilizadas. De modo a abranger todos os subeventos, pode acontecer que uma fotografia de pior qualidade seja mantida devido à pequena quantidade de fotografias de tal subevento. Desta forma, existe a possibilidade de que fotografias com melhor qualidade, mas pertencentes a subeventos com maior número de fotografias com qualidade superior, sejam preteridas.

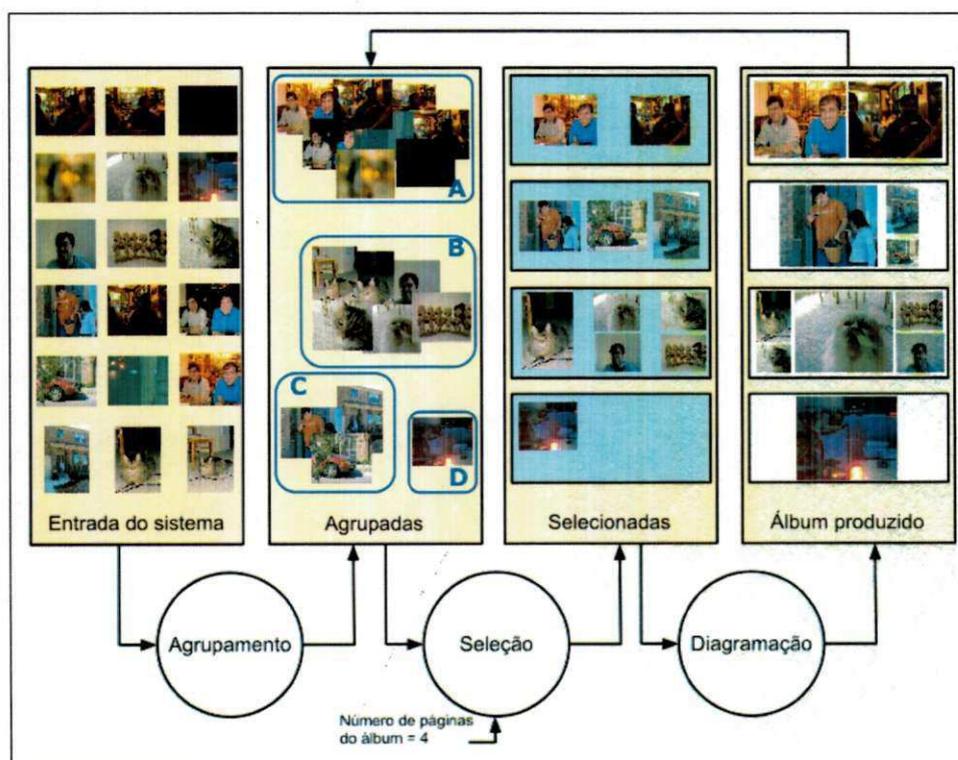
A partir das fotografias ordenadas por importância, o módulo gerador do álbum é executado. Além das fotografias, este módulo também requer as definições de características do documento alvo por parte do usuário, as quais são (i) as dimensões do álbum e (ii) o número desejado de páginas.

Para a produção do álbum, são necessárias duas etapas principais: (i) o agrupamento e a distribuição das fotografias ao longo das páginas do álbum e (ii) a diagramação das fotografias selecionadas em cada página.

Na primeira etapa, agrupamento, identificam-se similaridades existentes entre as fotografias, com o intuito de posicionar, nas mesmas páginas, aquelas que possuem algum tipo de vínculo semântico. Durante a etapa de agrupamento, não é levado em consideração apenas o resultado da ordenação, podendo-se eliminar, por exemplo, fotografias muito similares ainda que estas tenham sido bem classificadas pelo algoritmo ordenador.

Um diagrama esquemático do sistema proposto é ilustrado na Figura 3.2.

Figura 3.2: Ilustração do sistema proposto.



Na ilustração da Figura 3.2, a quantidade de grupos foi igual ao número de páginas definido pelo usuário e respeitou-se o número máximo de fotografias por página. Há, contudo, outras situações nas quais: (1) a quantidade de fotografias por página é maior do que o valor definido pelo usuário e/ou (2) a quantidade de grupos é maior

do que a quantidade de páginas. Em situações como estas, é necessário haver um novo processamento das páginas geradas durante a etapa de agrupamento. Na segunda iteração, busca-se equilibrar a quantidade de fotografias por página, minimizando-se, na medida do possível, a diferença entre as fotografias posicionadas em cada página. Este equilíbrio pode ser obtido pela redistribuição das fotografias ao longo das demais páginas do álbum, respeitando-se a ordenação previamente realizada.

Após concluída a etapa de agrupamento, inicia-se a segunda etapa da produção do álbum: a diagramação. Tal etapa é responsável pela definição do posicionamento e das dimensões de cada fotografia na página do documento em que tal fotografia será inserida. Tais etapas serão descritas em mais detalhes nos Capítulos 4, 5 e 6.

3.2 Formalização do problema

Nesta seção, busca-se descrever a abordagem proposta na seção anterior, utilizando-se de notação matemática.

Seja θ o conjunto de todas as fotografias digitais de um usuário e Ω um subconjunto de fotografias digitais de θ associado a um evento cujo início e fim possam ser identificáveis pelo usuário.

Deseja-se a obtenção de um conjunto Ω' , sendo $\Omega' \subset \Omega$, tal que $|\Omega'| = |\Omega| \times \rho$ e sendo ρ o percentual de fotos a serem escolhidas. Desta forma, torna-se necessária a definição de um procedimento f , a partir do qual seja possível selecionar, de acordo com critérios pré-definidos, as fotografias mais adequadas para compor o subconjunto Ω' . Portanto:

$$\Omega' = f(\Omega, \rho) \quad (3.1)$$

O procedimento f é o objeto de estudo do Capítulo 4.

O processo de diagramação de um álbum de fotografias consiste em, tendo-se Ω' e um álbum D composto de m páginas $\{P_1, P_2, \dots, P_m\}$, de dimensões pré-definidas (m deve ser definido pelo usuário do sistema), distribuir as n fotos, em que $n = |\Omega'|$, pelas m páginas de maneira que, em cada página, exista pelo menos uma foto, o que implica $n \geq m$.

A distribuição das n fotografias ao longo das m páginas se dá por meio de um

procedimento de distribuição g . Portanto, tem-se que:

$$\forall P_i \in D \exists \Omega_i'' | \Omega_i'' = g(i, \Omega'), \text{ em que } i \in \mathbb{Z}_+ \wedge i \leq m \wedge \Omega_1'' \cap \Omega_2'' \cap \dots \cap \Omega_n'' = \emptyset \quad (3.2)$$

O procedimento g é um dos objetos de estudo do Capítulo 5.

O passo final da diagramação é a distribuição das fotografias pertencentes a Ω_i'' na área disponível A da página P_i . Isto pode implicar a necessidade de recorte de bordas externas das fotografias, de forma a gerar proporções (*aspect ratios*) mais adequadas à diagramação. A área A de uma página depende de suas dimensões W_{pagina} e H_{pagina} (largura e altura, respectivamente) sendo $A = W_{pagina} \times H_{pagina}$. Seja l um procedimento que permite mapear as imagens pertencentes a Ω'' ao longo de uma região de P_i . O mapeamento da k -ésima imagem de Ω'' em P_i é definido pela 9-tupla de valores $(I_k, q_k, w_k, h_k, x_k, y_k, W_k, H_k, \alpha_k)$. Sendo assim:

$$l(\Omega_i'') = \{(I_k, q_k, w_k, h_k, x_k, y_k, W_k, H_k), \forall k \in \Omega_i''\} \quad (3.3)$$

Na tupla apresentada na Equação 3.3, I_k é a k -ésima imagem de Ω'' , q_k definem as coordenadas de um ponto da k -ésima imagem de onde se inicia o recorte, e w_k e h_k a largura e altura deste recorte, x_k e y_k são as coordenadas da página na qual será posicionada a k -ésima imagem, W_k e H_k são as dimensões as quais a k -ésima imagem ocupará na página, após o recorte. Por fim, α_k é o ângulo que a k -ésima imagem será rotacionada, apesar da rotação artística das fotografias estar fora do escopo desta pesquisa.

É possível definir algumas restrições sobre os valores que podem ser assumidos pelos elementos da 9-tupla resultante da aplicação do procedimento l mas, de forma geral, quaisquer combinações podem ser aceitas, desde que estejam de acordo com uma lógica de diagramação. As restrições, regras de otimização e avaliação serão posteriormente estudadas no Capítulo 6, o qual é dedicado ao estudo do procedimento l .

Em suma, o objetivo desta pesquisa é gerar um álbum D , formado por uma sequência de páginas P_i , conforme a Equação 3.4.

$$D = (P_1, P_2, \dots, P_m) | P_i = \{l(g(i, f(\Omega, \rho))), W, H, \dots\} \quad (3.4)$$

É relevante destacar que P_i é definido pela função l e por características tais como a cor do plano de fundo e o tipo da fonte a ser utilizada para a renderização da legenda da fotografia, dentre outras, que não estão sendo representadas na Equação 3.4 por questões de simplicidade, uma vez que não serão alvos de estudo nesta pesquisa. Outros elementos passíveis de estarem presentes em uma página, e.g., título, legendas, número da página, também foram suprimidos da Equação 3.4, por estarem fora do escopo desta pesquisa.

3.3 Considerações finais

Neste capítulo, foi apresentada a abordagem proposta para o problema investigado nesta pesquisa: diagramação automática de álbuns fotográficos. A descrição de alto nível ora proposta teve, como objetivo, apresentar a solução e permitir melhor compreensão da estrutura deste documento.

Quanto à descrição formal, esta foi realizada com o objetivo de desambiguar os detalhes da solução proposta, além de facilitar a compreensão da estruturação da solução, por intermédio da formulação matemática.

Com base no que foi apresentado e com o objetivo de sumariar parte do conhecimento sobre o tema investigado nesta pesquisa, apresenta-se o Quadro 3.1.

Quadro 3.1: Quadro comparativo entre as características das abordagens estudadas.

	Seleção de fotografias	Agrupamento	Leiaute	Intervenção do usuário
Obrador e Moroney (2009)	Medida do apelo	Cronologia	BRIC	Definição do número de fotografias
Xue et al. (2013)	não realizado	não realizado	BRIC	Definição do número de páginas
Egorova e Safonov (2013)	não realizado	<i>k-means</i> e cronologia	<i>Collage</i>	Seleção e agrupamento das fotografias
Bergmann et al. (2013)	Brilho, matiz, contraste e saliência	Detecção de duplicatas por cronologia e características GIST	BRIC	Seleção das fotografias
Sistema proposto	Regras de composição fotográfica	Cronologia e conteúdo	BRIC otimizado	Definição do número de páginas

Com base no Quadro 3.1, pode-se perceber que há uma ampla maioria de aborda-

gens que utilizam o algoritmo de diagramação BRIC. Também se percebe a opção pela utilização da cronologia na etapa de agrupamento, apesar de nenhum autor ter optado pela combinação da cronologia e do conteúdo da fotografia. Além disto, de acordo com o Quadro 3.1, percebe-se que a seleção das fotografias se dá nas abordagens anteriormente propostas, prioritariamente, a partir de medidas de baixo nível (e.g., brilho e apelo, que consiste em uma métrica definida pelos autores, a partir de características de baixo nível). Por fim, com relação à intervenção do usuário, percebe-se que, da mesma forma que o que foi proposto nesta pesquisa, existem outras abordagens que requerem do usuário pouca intervenção, a exemplo da abordagem proposta por Obrador e Moroney (2009), que requer apenas o número de páginas a ser utilizado na produção do álbum fotográfico.

A partir da estrutura ora apresentada neste capítulo, serão detalhados, nos capítulos subsequentes, cada um dos módulos supracitados.

Capítulo 4

Métodos para a análise de fotografias digitais

Durante a produção de um álbum fotográfico, o primeiro passo e, possivelmente, o mais importante é a seleção das fotografias, uma vez que estas são o objeto principal de exibição no documento a ser gerado.

A análise de fotografias é um campo de pesquisa que vem recebendo crescente atenção na literatura, tal e qual apresentado no Capítulo 2. O objetivo é, dada uma fotografia, classificá-la quanto à qualidade, seja a partir de uma avaliação numérica, seja por meio de um encadeamento de características extraídas. Vale enfatizar que o conceito de análise da qualidade de fotografias, no contexto desta pesquisa, não está relacionado a aspectos de baixo nível (e.g., resolução ou artefatos de compressão), mas ao conteúdo visual da fotografia e a sua atratividade.

Um grande desafio da análise automática de fotografias é realizar a análise de uma forma mais objetiva, evitando ambiguidades normalmente associadas a fatores subjetivos. A subjetividade normalmente está associada a fatores (e.g., os alvos humanos presentes, o local onde a fotografia foi obtida e o tipo de pose) que podem alterar a classificação das fotografias, em função do observador que realizará a avaliação.

Entretanto, é possível identificar fatores objetivos que viabilizam uma análise automática de fotografias. Neste sentido, no presente capítulo, são apresentados, propostos e avaliados métodos objetivos para a análise de fotografias, tendo como alvo fotografias obtidas por usuários amadores, ou seja, aquelas obtidas sem propósitos artísticos. O

escopo da análise engloba aspectos referentes à composição fotográfica, à harmonização cromática e à segmentação de regiões, conforme será detalhado a seguir, apesar de existirem diversos outros fatores que poderiam ter sido incluídos nesta investigação (e.g., análise do desfocamento de uma fotografia (LIU; LI; JIA, 2008b)).

Na Seção 4.1, é apresentado um novo método destinado à análise da composição fotográfica: o método para a avaliação do espaço negativo. Os métodos anteriormente descritos, i.e., aqueles adotados por Cavalcanti (2007), são também utilizados nesta pesquisa, dado que se trata de uma extensão da referida pesquisa. Neste caso, tais métodos são utilizados para analisar as fotografias e não modificá-las, conforme acontecia na pesquisa anterior.

Adicionalmente, na Seção 4.2, é apresentado um novo método para a análise da harmonia das cores presentes em uma imagem. Por fim, a Seção 4.4 contém as considerações finais deste capítulo.

4.1 Regras de composição fotográfica: proposta da regra do espaço negativo

As regras de composição fotográfica são aquelas que servem ao fotógrafo como guias de posicionamento do alvo na fotografia, de maneira a torná-la mais atrativa aos observadores. Neste sentido, uma forma de realizar a análise de fotografias é avaliar se uma dada fotografia está em conformidade com regras de composição fotográfica consideradas.

Todavia, tais regras possuem alto grau de subjetividade, sendo difíceis de serem traduzidas em algoritmos. Nesta seção, propõe-se a automatização de uma regra de composição até então não explorada na literatura: a regra do espaço negativo, cujo objetivo é avaliar se existe uma área de espaço negativo (ou seja, espaço não utilizado) aumentando o impacto da área preenchida. Tal regra identifica se há, em uma fotografia, espaços em que não exista informação visual relevante. O objetivo desta regra é produzir uma área de descanso ao olhar do observador, à medida que o direciona para o alvo da fotografia. De acordo com Berdan (2004), a regra do espaço negativo, também denominado espaço em branco, orienta a inclusão de áreas vazias em fotografias, as

quais constituem elementos visuais que conferem equilíbrio à fotografia.

Uma vez que a regra pode ser interpretada de diferentes formas, a regra do espaço negativo, conforme definido nesta pesquisa, considera uma fotografia (1) na orientação paisagem, (2) que possui o alvo não centralizado ocupando, preferencialmente, um dos terços da fotografia e (3) que tem o plano de fundo neutro (não possui elementos de distração).

A abordagem destinada ao cálculo de conformidade à regra do espaço negativo fundamenta-se na detecção das bordas entre as regiões que compõem a fotografia. A presença de bordas pode ser um indício de regiões complexas, enquanto regiões com poucas bordas podem ser consideradas regiões com espaço negativo. Existem, contudo, situações de alvos que apresentam poucas bordas (e.g., a fotografia de um ovo, em um cenário de uma única cor). Entretanto, o objetivo da realização da análise é a obtenção de informação sobre as fotografias, ainda que estas sejam aproximações grosseiras em algumas situações.

Para a detecção das bordas, foram avaliados os detectores *Canny*, *Laplace*, *Prewitt*, *Roberts*, *Sobel* e *Zerocross* (GONZALEZ; WOODS, 2001). Os melhores resultados, confirmados a partir de um processo de avaliação subjetiva, foram obtidos a partir da utilização do detector de *Canny*, tendo sido utilizados os parâmetros 24 (limiar inferior de corte) e 101 (limiar superior de corte), δ igual a 3 em máscara de tamanho 3x3, conforme descrito por Arruda (2009), uma vez que seu propósito foi enfatizar alvos humanos.

A partir das bordas extraídas da fotografia, contida no conjunto de fotografias adquirido, são obtidos histogramas de projeção horizontal e vertical, calculados a partir da soma das intensidades de todos os *pixels* pertencentes a uma coluna ou linha, respectivamente. De modo a facilitar a comparação entre fotografias de diferentes dimensões, histogramas de projeção médios foram obtidos com um número fixo de *bins* (100, nesta pesquisa). Histogramas de projeção médios podem ser definidos pela Equação 4.1.

$$H(I, i) = \sum_{x=\frac{(i-1) \times w}{b} + 1}^{\frac{(i-1) \times w}{b} + 3} \sum_{y=1} hI(x, y), \quad (4.1)$$

em que I é a matriz de intensidade da imagem, $I(x, y)$ é o valor de intensidade da

imagem no ponto de coordenadas (x, y) , w é a largura da imagem, h é a altura da imagem, i é o *bin* do histograma a ser calculado e b é o número de bins que compõem o histograma, sendo $b \leq w$.

Uma abordagem passível de ser adotada consiste na comparação dos histogramas de projeção de uma dada fotografia com os histogramas de projeção de um dado conjunto de fotografias previamente rotulado, seguido da avaliação de qual dos conjuntos apresenta maior semelhança com tal fotografia. Uma vez que foi definido que a regra do espaço negativo apenas será avaliada para imagens no formato paisagem, os histogramas de projeção a serem calculados a seguir são todos histogramas de projeção verticais.

Com o intuito de abranger ao máximo as diferenças existentes nas diversas situações que uma fotografia pode ser obtida, foi construída uma base de fotografias contendo tanto aquelas em conformidade com a regra do Espaço Negativo, quanto as demais não conformantes com a regra.

As fotografias selecionadas estão disponíveis em <https://sites.google.com/a/copin.ufcg.edu.br/negativespace/>. A seleção das imagens foi realizada por um voluntário, sendo levadas em consideração as três condições supramencionadas, de modo que a fotografia estivesse em conformidade com a regra do Espaço Negativo. Ao todo, 450 fotografias foram selecionadas de um conjunto de 2.778 imagens obtidas a partir de diversos bancos de imagens públicos, os quais são listados na Tabela 4.1.

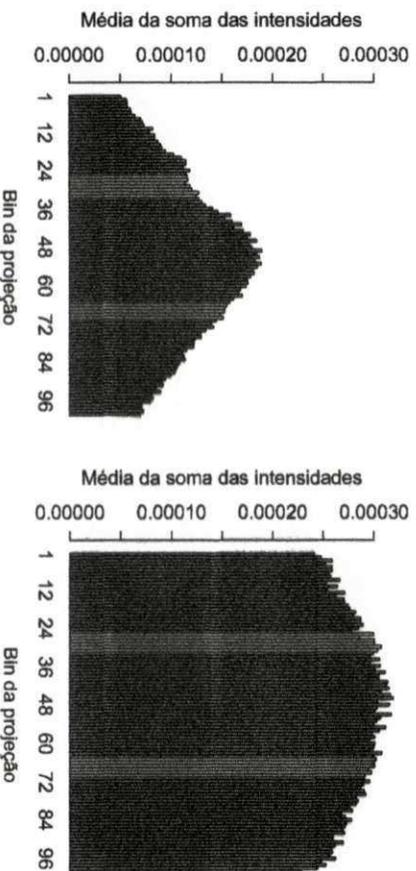
Tabela 4.1: Fontes de busca e número de imagens adquiridas por semente.

Semente de busca	Número de Imagens Adquiridas
MorgueFile (CONNORS; CONNORS, 2009)	500
Free-Stock-Images (Turbo photo inc., 2011)	410
StockVault (GUDMUNDSSON, 2011)	540
Free Photos Bank (FREE. . . , 2010)	670
Pixel Perfect Digital (HESTER, 2009)	658
Total	2778

Dois terços das imagens selecionadas (300 imagens) foram utilizadas para a produção do histograma de projeção médio, sendo metade destinado a cada classe, (1) em

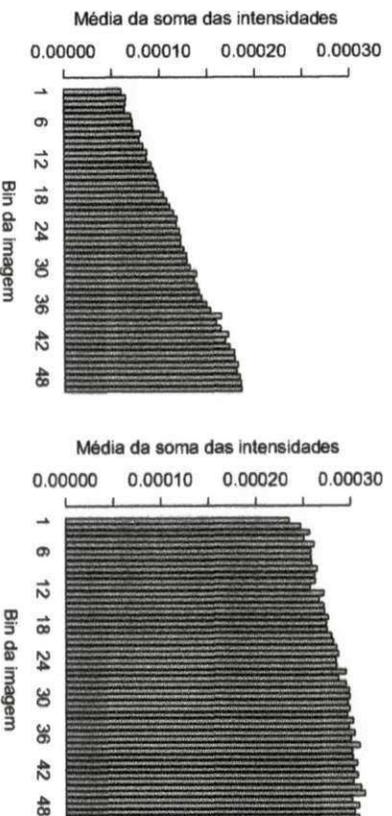
conformidade à regra do Espaço Negativo (2) não conformidade com a regra do Espaço Negativo. O histograma de projeção médio é apresentado na Figura 4.1.

Figura 4.1: Histogramas de projeção médios das fotografias que, respectivamente, obedecem à regra do Espaço Negativo e não obedecem à regra do Espaço Negativo.



A tendência de crescimento nas regiões de interesse da imagem pode ser evidenciada ao calcular o histograma de projeção da imagem de forma simétrica, das extremidades ao centro. Assim, tal histograma possui apenas 50 bins, referentes à média das colunas simetricamente somadas, ou seja, para o bin i , calcula-se a média entre todas as imagens nas colunas i e $100 - i$. Na Figura 4.2, ilustra-se o resultado deste histograma médio de projeção.

Figura 4.2: Histogramas de projeção médios e simétricos das fotografias que, respectivamente, obedecem à regra do Espaço Negativo e não obedecem à regra do Espaço Negativo.



Analisando o histograma de projeção médio das fotografias, pôde-se perceber uma

pequena tendência de crescimento nos histogramas de projeção daquelas que não estavam em concordância ao Espaço Negativo. Este comportamento já era esperado, pois havia uma tendência à centralização dos alvos em tais fotografias. Por sua vez, os histogramas de projeção de fotografias que obedecem à regra do Espaço Negativo apresentaram uma tendência de crescimento acentuada a partir do primeiro terço da fotografia e uma diferença significativa entre os maiores e menores valores computados.

Apesar de terem sido utilizados, prioritariamente, os histogramas de projeção vertical, analisou-se também o comportamento dos histogramas de projeção horizontal, nos quais não foi percebida diferença de comportamento. A partir da expansão da regra do espaço negativo, de modo a acomodar também fotografias com orientação vertical, possivelmente também podem vir a ser percebidas diferenças entre os histogramas anteriormente mencionados.

Tendo em vista que foram observadas diferenças no comportamento dos histogramas de ambas as classes, o passo seguinte consistiu em tentar obter uma relação que possibilitasse diferenciar numericamente tais classes. A primeira abordagem testada consistiu em calcular a diferença entre um histograma de projeção de fotografia em análise e o histograma médio de ambas as classes. A menor das diferenças foi inicialmente empregada para indicar um comportamento similar e, portanto, possibilitar a identificação da classe a qual o histograma pertencia. Para o cálculo do histograma médio, foram utilizadas 300 imagens, igualmente divididas entre conformantes e não conformantes à regra do Espaço Negativo.

As demais 150 imagens foram utilizadas para testar a abordagem. Obteve-se uma taxa de acerto média de 48%. Outras abordagens (e.g., *Earth mover's distance* (RUBNER; TOMASI; GUIBAS, 2000) e a comparação entre os histogramas a partir do chi-quadrado, definido pela Equação 7.5, foram testadas, tendo sido obtidos resultados similares.

Uma vez que a abordagem anteriormente testada não produziu bons resultados, foi testada uma abordagem que consistiu em utilizar os quartis do histograma q_1 , q_2 e q_3 , definidos, respectivamente, por $\frac{n+1}{4}$, $\frac{n+1}{2}$ e $\frac{3(n+1)}{4}$, estando os dados ordenados e sendo n o número de bins do histograma.

Assim, ao invés de se comparar diretamente os *bins* do histograma, foram com-

parados apenas os quartis de tais histogramas. Para tal abordagem, os resultados passaram a cerca de 55% de acerto médio, sendo uma melhora significativa em relação à abordagem anterior.

Entretanto, em ambas as abordagens percebeu-se que muitos erros ocorreram devido à comparação ser efetuada a partir dos valores absolutos dos *bins* do histograma e não a partir das distribuições dos mesmos.

De modo a evitar os problemas descritos anteriormente, foi testada uma abordagem utilizando Redes Neurais Artificiais (RNA) (HAYKIN, 1999). Decidiu-se pela utilização das RNA, em virtude da boa capacidade de generalização em problemas não-linearmente separáveis, como se configura o presente problema (HAYKIN, 1999). Pôde-se comprovar, conforme apresentado no Apêndice A.1, que a combinação das RNA com a PCA resultou em uma classificação correta de 86%, a partir da validação cruzada com 10 desdobramentos.

4.2 Análise da harmonia das cores

Conforme descrito na Seção 2.1.5, sobre Harmonização Cromática, as pesquisas existentes relacionadas à presente pesquisa tratam da harmonização automática de fotografias. Nesse caso, o principal objetivo é modificar as cores das fotografias, até que um determinado padrão de distribuição de cores seja obtido.

De modo a contornar as limitações ora apresentadas, foi proposta nesta pesquisa uma abordagem destinada à análise da harmonia de cores de uma fotografia (1) fundamentada em experimentos formais sobre combinações de cores, e (2) com a análise das cores da fotografia sendo realizada a partir das regiões de cor que representam regiões importantes da imagem.

Neste contexto, apresenta-se, nesta seção, um método destinado à segmentação por cor de modo a identificar regiões da fotografia.

4.2.1 Segmentação

Para que a análise das regiões de cor possa ser melhor realizada, é importante agrupar tais regiões. É imprescindível agrupar previamente cores que possam se confundir visu-

almente (e.g., todas as diferentes tonalidades que compõem a textura da pele humana). Entretanto, para o processo de análise das cores, não é necessário que todas as cores diferentes sejam segmentadas, pois regiões significativamente pequenas, na maioria dos casos, não influenciam a compreensão geral da imagem.

Para realizar essa segmentação, foram testados dois algoritmos: *Mean Shift* (COMANICIU; MEER, 2002) e o segmentador baseado em grafos (FELZENSZWALB; HUTTENLOCHER, 2004). Estes algoritmos foram selecionados por (i) serem métodos amplamente utilizados com esta finalidade (ZHANG; FRITTS; GOLDMAN, 2008); (ii) possuírem implementação gratuita e disponível; e (iii) produzirem resultados equivalentes aos demais (PANTOFARU; HEBERT, 2005; ESTRADA; JEPSON, 2009; ALBATAL et al., 2009) (o que pode também ser comprovado pelo fato de não existir, de acordo com a pesquisa realizada, uma métrica que garanta a superioridade de um método com relação aos demais (ZHANG; FRITTS; GOLDMAN, 2008));

A seguinte abordagem foi testada com o objetivo de detectar os melhores parâmetros de segmentação: foi construída uma base de dados contendo imagens de referência (*groundtruth*), segmentadas manualmente, a fim de representarem o resultado esperado; utilizando-se algoritmos genéticos e se dispondo dos genomas correspondentes a uma combinação de parâmetros para os segmentadores, foram obtidas imagens automaticamente segmentadas, a partir das imagens originais; a diferença D obtida da comparação entre as imagens de dimensões $m \times n$ geradas automaticamente e manualmente (apresentada na Equação 4.2), foi utilizada como medição da adequação do genoma ao problema, sendo os melhores genomas recombinaados para a obtenção da próxima geração.

$$D = \sum_{i=1}^m \sum_{j=1}^n e(P_1(i, j), P_2(i, j)), \quad (4.2)$$

em que $P_1(x, y)$ e $P_2(x, y)$ representam a cor do pixel de coordenadas (x, y) das imagens 1 e 2 respectivamente e a função $e(a, b)$ calcula a distância euclidiana entre os pontos a e b

À medida que o treinamento supradescrito era executado, pôde-se perceber alguns possíveis melhoramentos (pré ou pós processamentos) que poderiam auxiliar o algoritmo de segmentação na obtenção de melhores resultados. Portanto, o genoma passou a conter também parâmetros para os algoritmos de processamento utilizados (e.g., pa-

râmetros para o filtro de suavização). Entretanto, a ordem de execução dos elementos de processamento ainda era pré-definida. Esta limitação poderia ocultar possíveis soluções, até então impensadas. Assim, uma forma de ampliar o potencial de busca por soluções seria a possibilidade de se alterar a ordem de execução dos elementos de processamento. Para que esta variação fosse realizada de forma metodológica, foi desenvolvido um método evolucionário, o qual é descrito no Apêndice B.

Uma vez segmentada a imagem, o passo seguinte foi identificar quais das regiões possuíam maior importância para a imagem. Esta identificação deu-se a partir do cálculo das regiões da imagem, por intermédio do algoritmo de componentes conectados. Obteve-se como resposta um grafo no qual cada nó corresponde aos *pixels* pertencentes a uma mesma região, enquanto cada vértice corresponde à ligação entre regiões vizinhas.

Para a escolha dos nós mais significativos (e, por conseguinte, das cores mais significativas), duas abordagens foram testadas: (1) os nós compostos pelas regiões com maior número de *pixels* e (2) os nós com maior número de vértices. Ambas as abordagens visaram a selecionar as regiões de maior área.

A quantidade de nós pode tanto ser escolhida pela análise da quantidade de nós encontrados e o contraste existente entre as cores e ligações entre tais nós quanto ser previamente definida. Uma vez que os métodos utilizados nesta pesquisa para a comparação apresentada na Seção 7.2.2 (i.e., (INC., 2013a; YAO et al., 2011)) e aqueles tipicamente utilizados para esta tarefa (e.g., o *k-means* (BISHOP, 2006)) requereram prévia definição do número de cores, decidiu-se pela utilização do método em que o número de nós seria previamente definido. Por outro lado, em virtude dos métodos comparados utilizarem números diferentes de cores, definiu-se 5 como o número de cores adotado no processo de comparação com a ferramenta Kuler (INC., 2013a) e 3 como o número de cores adotado no processo de comparação com a abordagem de Yao et al. (2011). Dado que o *k-means* é flexível quanto ao número de cores, a comparação com tal método foi feita tanto com 3 quanto com 5 nós (ou cores). A avaliação dos resultados é descrita em mais detalhes na Seção 7.2.2.

4.2.2 Cálculo da harmonia

Uma vez selecionadas as k cores para representar a imagem, o passo seguinte foi avaliar a harmonia entre essas cores, o que foi realizado a partir do método descrito por Sauvaget et al. (2010), o qual é inspirado na tabela de proporção de áreas de contraste proposta por Itten (1961). A tabela de proporção de áreas de contraste é explicada em mais detalhes na Seção 2.1.5.

Uma vez calculada a tabela de proporções de cores para as k cores selecionadas, calcula-se a diferença (*erro*) entre as proporções encontradas e as proporções ideais, definidas por Sauvaget et al. (2010) da seguinte forma:

$$erro = \sum_{S \in CS} |ideal(S) - tamanho(S)|, \quad (4.3)$$

em que CS são os setores de cor definidos por Itten (1961), S é cada um dos setores de CS , $ideal(S)$ é a proporção ideal de *pixels* que o setor S deveria apresentar (de acordo com a tabela de proporção de áreas de contraste) e $tamanho(S)$ é a proporção de *pixels* da imagem cuja tonalidade pertence ao setor S .

4.3 Combinação dos avaliadores

Uma vez descritos os dois métodos propostos nesta pesquisa, faz-se necessária a descrição de como tais métodos foram combinados aos métodos já existentes (descritos na Seção 2.1) para a realização da análise de uma fotografia. Foram utilizadas as métricas: intensidade média dos pixels, saturação média dos pixels, matiz média dos pixels, matiz da região dos terços, saturação da região dos terços, intensidade da região dos terços, as dimensões da imagem, o *aspect ratio* (DATTA et al., 2006) e o contraste (KE; TANG; JING, 2006), uma vez que estas métricas foram consideradas as mais adequadas nas respectivas pesquisas.

Além destas métricas, também foram utilizadas as seguintes métricas:

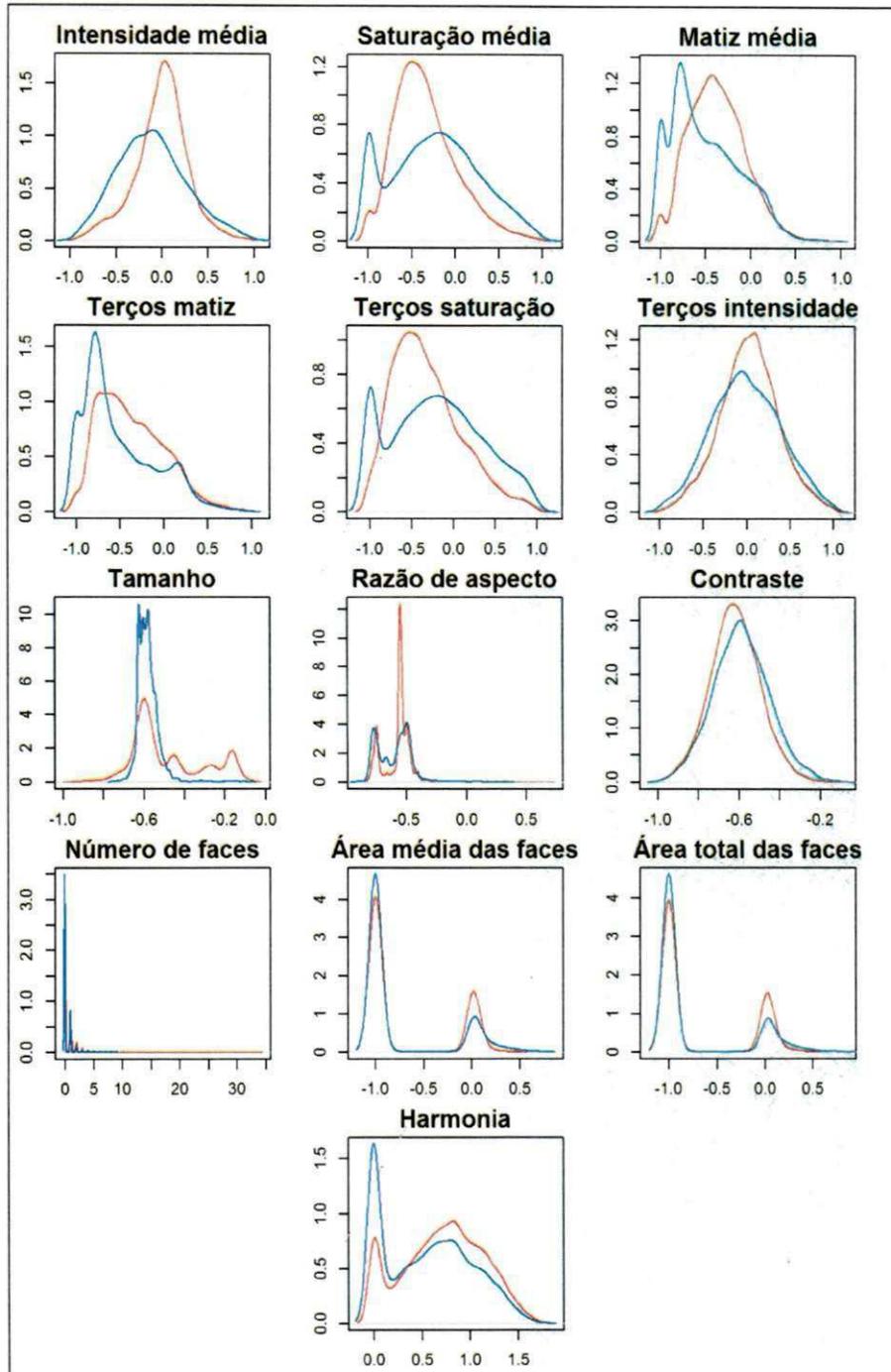
- **O número de faces detectadas.** Definido por um número inteiro maior ou igual a zero, representando o número de faces detectadas pelo detector de faces utilizado nesta pesquisa, i.e., o detector concebido por Viola e Jones (2001);

- **A área média das n faces detectadas.** Definida por $\frac{1}{n} \times \sum_{j=1}^n \frac{A_{f_j}}{A_i}$, em que A_{f_j} é a área da face j e A_i é a área total da imagem i ;
- **A área total das n faces detectadas.** Definida por $\frac{\sum_{j=1}^n A_{f_j}}{A_i}$, em que A_{f_j} é a área da face j e A_i é a área total da imagem i ; e
- **A harmonia das cores.** Definida na Seção 4.2.2.

Para cada um dos métodos descritos, calcularam-se os valores com base em um conjunto de fotografias previamente rotuladas como alta ou baixa qualidade, a partir de uma rotulagem realizada por voluntários em que uma fotografia era considerada de uma determinada classe se 10 dos participantes concordassem com a atribuição (LUO; WANG; TANG, 2011). Tais valores foram normalizados para o intervalo $[-1,1]$, uma vez que este intervalo favorece alguns algoritmos de aprendizagem (e.g., RNA) à medida que tal normalização evita que medidas realizadas em faixas de valores diferentes impactem na decisão do algoritmo de classificação utilizado. A normalização também poderia ter sido realizada para o intervalo $[0,1]$. Para as características que não apresentaram limiar superior (i.e., número de faces e o cálculo de harmonia), a normalização não foi realizada. Foram gerados gráficos da função da densidade de probabilidade por métrica nos quais a distribuição das fotografias consideradas com alta e baixa qualidade são traçadas na cores azuis e vermelhas, respectivamente. Tais gráficos são exibidos na Figura 4.3.

Pode-se perceber, a partir da avaliação destes dados, que as distribuições de ambas as classes se assemelham, o que sugere que, para o conjunto utilizado e para as métricas revisadas, não deve haver uma separabilidade linear entre as classes de fotografia. Tal resultado era esperado uma vez que o julgamento da qualidade da fotografia não se dá apenas por métodos mensuráveis mas também pela avaliação subjetiva, o que envolve, por exemplo, questões culturais e o objetivo do fotógrafo. Em algumas ocasiões, o contexto da fotografia pode ter levado o(s) participante(s) a considerá-la de baixa qualidade, mas a qualidade visual da mesma pode ser considerada, para outro conjunto de participantes, de alta qualidade.

Figura 4.3: Densidade de probabilidade das fotografias consideradas com alta qualidade (em azul) e das consideradas de baixa qualidade (em vermelho) em cada uma das métricas utilizadas.



Esta inconsistência dificulta a obtenção de um padrão de comportamento que relacione as métricas utilizadas e o julgamento dos participantes. De modo a comprovar tal

afirmação, foram testados diversos algoritmos (e.g., regressão linear (BISHOP, 2006), redes neurais artificiais (HAYKIN, 1999) e árvores de decisão (BISHOP, 2006)), com o intuito de aprender ou identificar uma correlação entre as avaliações numéricas realizadas e a rotulagem humana.

Não foi possível, mesmo realizando-se diversas variações paramétricas entre os métodos, encontrar convergência no treinamento dos métodos utilizados (i.e., resultados não foram significativamente melhores que 50%).

Em decorrência deste resultado, decidiu-se criar um algoritmo destinado a servir como estimativa de qualidade sem, contudo, realizar uma categorização das imagens. Desta forma, uma vez que cada métrica avalia um determinado item de qualidade, nesta abordagem decidiu-se avaliar a quantidade de métricas nas quais uma determinada fotografia violou os limites de qualidade pré-definidos. Os valores limítrofes para a avaliação da qualidade foram definidos de acordo com o conhecimento das métricas analisadas.

Os resultados do experimento realizado permitiram concluir que a combinação das métricas de seleção com o agrupamento realizado por tempo e similaridade produzem seleções similares àquelas realizadas por humanos, conforme apresentado na Seção 7.4.1.

4.4 Considerações finais

Neste capítulo, foram apresentados métodos destinados à análise de fotografias. Inicialmente, discutiu-se a utilização de algoritmos para a análise da conformidade de uma dada fotografia às regras de composição fotográfica. Em seguida, descreveu-se uma abordagem para a avaliação da regra do espaço negativo. Por fim, apresentou-se a proposta para a avaliação da harmonia cromática de fotografias, no âmbito da qual foi desenvolvido um segmentador de fotografias destinado a melhorar o processo de obtenção das regiões de cores mais importantes.

As regras de composição fotográfica podem ser utilizadas para a análise qualitativa das fotografias. Neste capítulo, foi descrito um método para avaliar a regra do espaço negativo. Entretanto, o uso de regras isoladas não é suficiente para qualificar uma fotografia, uma vez que a conformidade às regras não garante uma fotografia de qualidade.

Almeja-se, em investigações futuras, aumentar o número de regras de composição e avaliar quais possuem maior correlação com o julgamento humano.

Capítulo 5

Agrupamento de fotografias

Uma vez analisadas e selecionadas as fotografias, a etapa seguinte consiste na diagramação do álbum fotográfico. Nos casos em que o número de páginas é igual ao número de fotografias, não há a necessidade de se realizar nenhum processamento extra, uma vez que cada fotografia ocupará uma página. Para decidir quais as fotografias que deverão ser posicionadas próximas entre si, podem ser utilizados métodos que avaliem o grau de relacionamento das fotografias. Este é o objeto de estudo do presente capítulo.

Existem diversas formas de distribuir fotografias ao longo das páginas de um álbum fotográfico. Pode-se destacar duas formas comumente utilizadas: (1) aleatoriamente e (2) a partir de um agrupamento realizado por intermédio de uma função de similaridade. No posicionamento aleatório, as fotografias são distribuídas ao longo das páginas do álbum de forma arbitrária. Por outro lado, no posicionamento com base em função de similaridade, é necessário que seja definida uma função para comparar todas as fotografias duas a duas. O resultado da análise de similaridade é, então, utilizado para decidir as páginas em que cada fotografia deverá ser posicionada.

Pode-se elencar as seguintes vantagens associadas a cada um dos dois métodos apresentados anteriormente:

(1) Método aleatório:

- por ser computacionalmente simples, confere rapidez ao processo de geração do álbum;
- reduz a possibilidade de fotografias similares aparecerem próximas entre si; e

- não impõe pré-requisitos de funcionamento (e.g., existência de metadados contendo data e hora em que a fotografia foi obtida);

(2) Método fundamentado na análise de similaridade:

- permite que o histórico do evento seja acompanhado de forma mais eficaz; e
- evita que fotografias sem relacionamento algum entre si sejam consideradas similares, à medida que podem haver fotografias com cronologia similar, mas conteúdos diferentes.

Nesta pesquisa, optou-se pela utilização de uma função de análise da similaridade. Tal decisão foi tomada partindo-se da premissa de que, na criação de um álbum a partir das fotografias de um evento, afigura-se conveniente para manter a consistência entre as fotografias. Sendo assim, na Seção 5.1 é apresentada a função de similaridade proposta e adotada nesta pesquisa para o agrupamento de fotografias. Na Seção 5.2, é apresentado o método iterativo desenvolvido para a distribuição das fotografias ao longo do álbum fotográfico. Por fim, na Seção 5.3, são apresentadas as considerações finais.

5.1 Função de similaridade proposta

Não há, *a priori*, uma metodologia pré-estabelecida para a organização de fotografias, a qual, comumente pode ser realizada a partir de três fatores primordiais: o tempo, o local e o conteúdo (NAAMAN et al., 2004; RODDEN; WOOD, 2003). A decisão sobre qual fator levar em consideração é meramente individual. Enquanto alguns preferem a organização orientada ao conteúdo e ao local (e.g., todas as fotos contendo a torre Eiffel, independentemente da data e hora em que foram obtidas), outros preferem a organização cronológica (a qual respeita prioritariamente a ordem de obtenção das fotografias).

Nesta pesquisa, a organização fundamenta-se (mas não se limita) na cronologia, uma vez que se trata de uma estratégia de fácil assimilação e possui uma lógica de mensuração. Tal organização é majoritariamente utilizada pelos fotógrafos (profissionais ou amadores) para organizar suas fotografias (GOZALI; KAN; SUNDARAM,

2012). Para se obter melhores resultados, uma vez determinadas as fotografias similares cronologicamente, realiza-se uma análise do conteúdo para se identificar duplicatas. Este segundo processo é realizado após a conclusão do primeiro, dado que dificilmente ocorrerão duplicatas em períodos muito distantes de tempo.

Esta seção divide-se nas subseções 5.1.1 e 5.1.2, nas quais descrevem-se as estratégias de análise cronológica e do conteúdo respectivamente. Tais análises combinadas consistem no método utilizado nesta pesquisa.

5.1.1 Similaridade cronológica

Para que seja realizada a organização pela cronologia, faz-se mister que alguns requisitos sejam atendidos, em especial os seguintes:

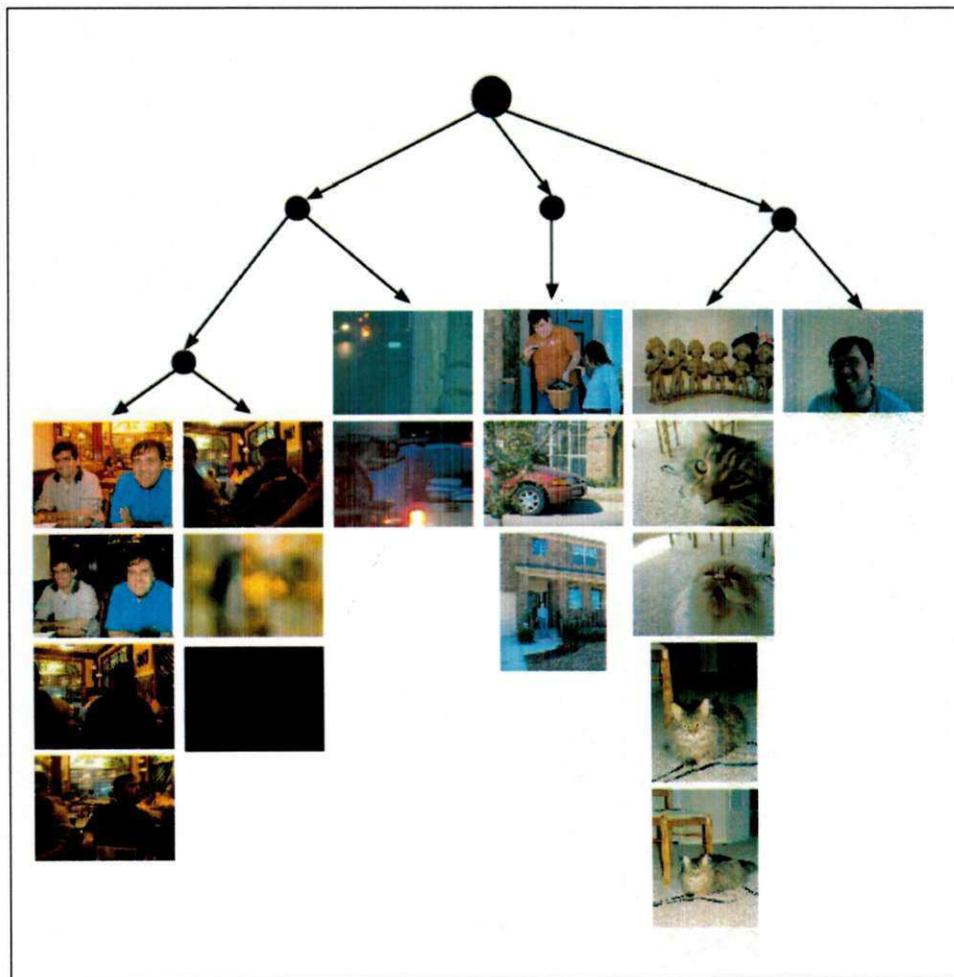
- Todas as fotografias devem ter sido obtidas em um evento delimitado pelo tempo compreendido entre a primeira e a última fotografia;
- um mesmo equipamento fotográfico deve ter sido empregado para a aquisição das fotografias, ao longo do evento ou os equipamentos utilizados estão com os seus relógios sincronizados;
- o equipamento fotográfico utilizado deve registrar as informações de tempo nos metadados da fotografia; e
- o relógio do equipamento deve estar ajustado e não deve ter sido modificado ao longo do evento;

Tais requisitos se fazem necessários para garantir que a cronologia do evento não será alterada e que todas as fotografias existentes são pertencentes ao mesmo evento, de modo a poderem compor o mesmo álbum fotográfico.

Na Seção 2.1.7, discutiu-se métodos para o agrupamento de fotografias. Em particular, foi apresentado o método proposto por Obrador e Moroney (2009). Dentre as desvantagens da abordagem utilizada por Obrador e Moroney (2009), podem-se citar (1) a necessidade de se definir o critério de parada da recursividade, (2) a possível sobresegmentação do evento e, por fim, (3) a divisão binária do tempo, o que pode impedir que um nó represente um único sub-evento.

Por conta das desvantagens apresentadas anteriormente, decidiu-se por um modelo alternativo. As principais diferenças deste modelo alternativo proposto para aquele utilizado por Obrador e Moroney (2009) são: (1) a divisão dos nós em mais de dois filhos e (2) a definição de limiares para os intervalos de tempo que determinem o grau de diferença temporal entre duas fotografias consecutivas. Esta opção permite que a estrutura dos subeventos seja obtida de forma mais eficaz à medida que estes são definidos como folhas de um mesmo nó. Tal estruturação permite, por exemplo, que as fotografias pertencentes a subeventos relacionados a um mesmo evento, que sejam espaçadas temporalmente mas que possam conteúdo similar, possam ser posicionadas próximas entre si. O modelo de organização hierárquica proposto é ilustrado na Figura 5.1.

Figura 5.1: Modelo proposto nesta tese de organização hierárquica pelo tempo.



Os limiares temporais anteriormente mencionados são valores pré-definidos com o intuito de descrever o intervalo de tempo entre sub-eventos. Estes limiares são utilizados pois não há um intervalo único que se adeque a todas as possíveis situações, e.g., para um evento que durou uma semana, pode ser razoável ter sub-eventos espaçados por um dia, enquanto para cuja duração foi dois dias, tal espaçamento pode não ter a mesma granularidade temporal.

De modo a reduzir tal problema, nesta pesquisa, foram definidos alguns valores: 30 segundos, 5 minutos, 30 minutos, 1 hora, 8 horas, 1 dia, os quais servem para representar diversos graus de espaçamentos temporais em eventos, e.g., deslocamento de um ponto turístico para outro, ajuste do equipamento e mudança de um dia para um outro. Assim, eventos separados por um espaço de tempo de 5 minutos tendem a ser mais relacionados do que eventos separados por espaços de tempo superiores a 1 hora.

Para a obtenção deste modelo hierárquico, utiliza-se o método representado pelo Algoritmo 1:

Algoritmo 1 Obtenção da árvore hierárquica de tempo.

```

1: ENCONTRARGRUPOS(S)
   //Inicializa uma variável auxiliar identificadora do início do grupo.
2:    $p \leftarrow 1$ 
   //Para cada fotografia do conjunto, calcular a diferença entre os horários de imagens consecutivas.
3:   para  $i \leftarrow 2$  até  $n$  faça
4:      $diff \leftarrow t(S_i) - t(S_{i-1})$ 
   //Para cada limiar de tempo,
5:     para  $j \leftarrow 1$  até  $l$  faça
   //verificar se a diferença entre as fotografias excede o limiar
6:       se  $diff \geq L_j$  então
   //Em caso afirmativo, adicionar a sequência ao conjunto,
7:          $M \leftarrow M \cup \langle p, i - 1 \rangle$ 
   //e refinar a busca utilizando limiares de tempo mais restritos.
8:          $M \leftarrow M \cup \text{ENCONTRARGRUPOS}(\{S_p, S_{p+1}, \dots, S_{i-1}\})$ 
   //Atualizar o valor de p com o índice do último grupo encontrado
9:          $p \leftarrow i$ 
10:      fim se
11:    fim para
12:  fim para
13:  devolve M
14: fim

```

em que n representa o número de fotografias da sequência de fotografias S (em que cada fotografia é representada por um índice S_i), L o conjunto com os l limiares de tempo previamente definidos e M um conjunto de pares ordenados que indicam os índices das fotografias iniciais e t é uma função destinada à conversão do horário da fotografia em um número inteiro. No Algoritmo 1, primeiro identifica-se, na sequência de fotografias, quais os pares adjacentes que possuem diferenças temporais superiores a um dos limiares pré-definidos, cujo conjunto é representado por L . À medida em que se encontra um par com tal característica, utiliza-se o conjunto delimitado pelas fotografias iniciais e finais pertencentes ao grupo. Uma vez obtidos os subconjuntos pertencentes àquele limiar temporal, calcula-se, para cada um deles, recursivamente, uma nova divisão dos conjuntos utilizando-se um limiar temporal menor. O algoritmo é repetido até que cessem as divisões ou os limiares.

5.1.2 Similaridade por conteúdo

A identificação de fotografias similares também pode se dar por seu conteúdo. Tal identificação é útil nos casos em que não há informação sobre o horário no qual as fotografias foram obtidas, assim como serve para a detecção de duplicatas, conforme verificado nos experimentos realizados nesta pesquisa. Entretanto, foi dada maior ênfase à detecção de duplicatas. As duplicatas, no contexto da construção de um álbum fotográfico, precisam ser descartadas, uma vez que não trazem informações novas à descrição visual do evento.

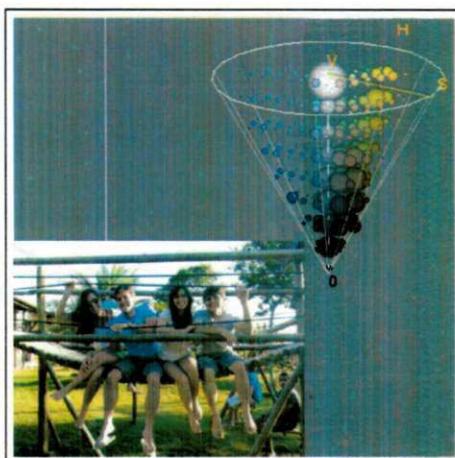
Nesta pesquisa, duplicatas foram definidas como fotografias que:

- pouco diferem nos ângulos de rotação e na translação;
- não possuem alterações quanto ao número e posicionamento de indivíduos e objetos;
- possuem os mesmos indivíduos nas mesmas posições e expressões faciais similares;
- e
- foram obtidas em um mesmo sub-evento.

A partir destes critérios, decidiu-se pelo desenvolvimento de um módulo de detecção de duplicatas que apresentasse boas taxas de detecção, porém que fosse realizado de forma rápida, uma vez que há a necessidade de se realizar comparações par a par entre todas as fotografias do conjunto. Assim, caso k fotografias tenham sido detectadas como pertencentes a um sub-evento, são necessárias $\frac{k \times (k-1)}{2}$ comparações.

Inspirado nos métodos discutidos no Capítulo 2 (fundamentação teórica) (WU et al., 2011), foi desenvolvido, para este propósito, um algoritmo destinado à análise de similaridade por cor, utilizando-se o histograma tridimensional do espaço de cor HSV e dividindo-se a imagem em regiões. O histograma 3D de uma imagem é ilustrado na Figura 5.2.

Figura 5.2: Histograma tridimensional de uma imagem.



A medida de similaridade fundamenta-se no cálculo de histogramas de cor em três dimensões. Os histogramas são obtidos a partir de n regiões da imagem, resultantes da divisão da imagem em v segmentos verticais e h segmentos horizontais, tendo-se $n = h \times v$. Esta divisão é feita para evitar que fotografias simétricas sejam consideradas duplicatas e evitar que fotografias diferentes mas com distribuições similares de cores sejam consideradas similares. Os valores típicos de h e v utilizados nesta tese foram iguais a 3 para ambas as variáveis.

Assim, para cada par de imagens, tem-se n histogramas de cor a serem comparados. São realizadas n comparações, já que cada histograma é comparado apenas com um

outro histograma na mesma posição. Esta abordagem tem a vantagem de acelerar o processo, uma vez que não se buscarão por outras regiões. A principal desvantagem desta abordagem é quando da ocorrência de rotação e translação. Assim, a definição da quantidade/do grau de translação e rotação permitidos nas fotografias para que estas possam ser consideradas similares, conforme destacado anteriormente, pode ser relacionada ao número de regiões existentes.

A abordagem utilizada requer a definição de alguns parâmetros, a saber: (1) o número de regiões verticais e horizontais nas quais a imagem será dividida, (2) o número de *bins* do histograma para cada componente do espaço e, por fim, (3) o método a ser utilizado na comparação dos histogramas. Na Seção 7.3, são apresentados os experimentos realizados para o cálculo da melhor combinação de parâmetros.

Uma vez identificadas as duplicatas, deve-se optar por sua exclusão ou não. Uma vez escolhida a exclusão, foi utilizado o método descrito no Capítulo 4 para ordenar e remover fotografias com aspecto visual inferior àquelas que serão selecionadas para compor o álbum.

Melhores taxas de detecção de duplicatas poderiam ter sido encontradas, caso tivessem sido utilizados métodos analisadores de expressões faciais e da quantidade de alvos humanos envolvidos nas fotografias, dado que estes aspectos são normalmente levados em consideração nas definições de duplicatas. Alterações na fotografia, ainda que sutis (e.g., pequenas alterações nas expressões faciais dos alvos humanos), podem dificultar a rotulagem de duplicatas (e.g., os mesmos alvos humanos fotografados, apresentando as mesmas poses, porém, diferentes estados, com relação à abertura dos olhos ou com relação à fotogenia).

5.2 Distribuição das fotografias ao longo das páginas do álbum fotográfico

Ao final do agrupamento hierárquico das fotografias (Seção 5.1.1) e da detecção e exclusão das fotografias duplicatas (Seção 5.1.2), a etapa final do processo de agrupamento

é a distribuição das fotografias selecionadas ao longo das páginas do álbum fotográfico.

No processo de distribuição das fotografias selecionadas, utiliza-se o resultado do agrupamento como o ponto de partida. Uma vez que a definição do número de páginas e do número máximo de fotografias por página são requisitos básicos do sistema proposto, e considerando que foi satisfeito o requisito do número mínimo de fotografias (igual ao número de páginas que o álbum terá), seis situações são possíveis, i.e., há:

1. um número de grupos igual ao número de páginas e nenhum grupo possui mais fotografias do que o número máximo permitido;
2. um número de grupos igual ao número de páginas e algum grupo possui mais fotografias do que o número máximo permitido;
3. mais grupos do que páginas e nenhum grupo possui mais fotografias do que o número máximo permitido;
4. mais grupos do que páginas e algum grupo possui mais fotografias do que o número máximo permitido;
5. menos grupos do que páginas e nenhum grupo possui mais fotografias do que o número máximo permitido;
6. menos grupos do que páginas e algum grupo possui mais fotografias do que o número máximo permitido;

O primeiro caso é o mais simples de ser resolvido: cada grupo ocupará uma página do álbum. Nos demais casos, existe a necessidade de remanejamento de fotografias até que existam tantos grupos quanto páginas. Para o caso onde existem mais grupos que páginas, foi desenvolvido um método iterativo, segundo o qual, a partir do cálculo da diferença temporal entre fotografias, pode-se ordená-las em função dos limiares definidos, de modo a permitir a identificação dos agrupamentos. Tal método iterativo é representado no Algoritmo 2.

Algoritmo 2 Remanejamento de fotografias ao longo dos grupos.

```

1: REMANEJAMENTO(G)
   //Para cada uma das fotografias em G
2:  para  $i \leftarrow 1$  até  $n(G)$  faça
   //Avaliar quantos grupos possuem mais fotografias que o máximo permitido
3:   se então  $n(g_i) > MAX$ 
   //Calcular quantos novos grupos serão criados
4:    $w \leftarrow \frac{n(g_i)}{MAX}$ 
5:   para  $j \leftarrow 2$  até  $w$  faça
   //Criar novo grupo com fotografias excedentes
6:    $l \leftarrow \{S_{j \times MAX}, \dots, S_{j \times MAX + MAX}\}$ 
7:   fim para
8:   fim se
9: fim para
   //Em havendo um número de grupos maior que MAXPAG
10: enquanto  $n(G) > MAXPAG$  faça
   //Para cada grupo existente,
11:   para  $p \leftarrow 2$  até  $n(G)$  faça
   //Calcular o número de fotos necessárias para que as duas páginas estejam preenchidas e
12:    $d \leftarrow (MAXPAG - n(g_p)) + (MAXPAG - n(g_{p-1}))$ 
   //Calcular a diferença de tempo entre ambos os grupos de fotografias.
13:    $d_t \leftarrow t(g_p) - t(g_{p-1})$ 
   //Armazenar as diferenças de espaço
14:    $P \leftarrow P \cup d$ 
   //e armazenar as diferenças de tempo
15:    $T \leftarrow T \cup d_t$ 
16:    $Q \leftarrow Q \cup \langle g_p, P, T \rangle$ 
17:   fim para
   //Ordenar o conjunto decrescentemente pelas diferenças de espaço
18:    $Q \leftarrow o(Q, P)$ 
   //Selecionar os 10% das diferenças com mais espaço disponível
19:    $R \leftarrow \{Q_1, Q_2, \dots, Q_{0,1 \times |Q|}\}$ 
   //Ordenar os grupos crescentemente pela diferença de tempo entre suas fotografias
20:    $R \leftarrow o(R, T)$ 
   //Concatenar as duas páginas com menor diferença temporal
21:    $G \leftarrow c(G_1, G_2)$ 
22: fim enquanto
23: fim

```

em que G é o conjunto de grupos, g_p é o p -ésimo grupo, n é uma função que indica o número de fotografias de uma página ou grupo, MAX é o número máximo de fotografias por página, $MAXPAG$ é o número máximo de páginas que o álbum poderá possuir, t é uma função destinada à conversão do horário da fotografia em um número inteiro e o é uma função destinada à ordenação de um conjunto a partir do segundo argumento.

Este algoritmo foi desenvolvido com o intuito de balancear o número de fotografias por página, evitando que este número ultrapasse o valor máximo estipulado pelo usuário.

Para os demais casos, nos quais existem menos grupos que páginas, é preciso quebrar algum agrupamento previamente realizado. O algoritmo para realizar tal tarefa deve identificar os agrupamentos que possuem mais fotografias e escolher uma delas (e.g., a partir do método para análise da fotografia) para a criação de um novo grupo.

Os algoritmos supramencionados foram assim desenvolvidos, de modo a integrar simplicidade e velocidade. Entretanto, outras abordagens são possíveis, e.g., a otimização multiobjetivos.

5.3 Considerações finais

Neste capítulo, foi apresentada a proposta de um método para o agrupamento das fotografias de uma coleção arbitrária. Tal agrupamento é essencial para que sejam removidas as duplicatas e para que o agrupamento seja realizado corretamente.

O método proposto utiliza as informações de tempo e as distribuições de cor para identificar as duplicatas. A partir de uma avaliação experimental, pôde-se constatar que o método presta-se satisfatoriamente aos propósitos desta pesquisa. Mais detalhes do processo são apresentados no Capítulo 7.

Capítulo 6

Método para a diagramação automática de álbuns fotográficos

A etapa final da abordagem proposta nesta tese é a produção do álbum fotográfico. Nesta etapa, objetiva-se dispor todas as fotografias previamente escolhidas e ordenadas por importância em um único documento.

Desta forma, neste capítulo é descrito um método para a diagramação de álbuns fotográficos que considere, na medida do possível, as restrições definidas na pesquisa de Atkins (2008). Tais restrições estão transcritas na Seção 2.2, dividem-se em restrições primárias e secundárias e dizem respeito a quais princípios de diagramação devem ser priorizados em detrimento dos demais (e.g., a principal restrição é manter a relação de aspecto, sendo assim uma restrição de menor grau de importância não pode forçar a perda da relação de aspecto).

Existem diversas formas de se realizar tal distribuição espacial das fotografias. Na prática, uma disposição aleatória é uma abordagem que pode agradar a alguns usuários. Desta forma, são inúmeras as possibilidades de diagramação de um álbum fotográfico. Identificar qual o estilo de diagramação que mais agrada ao cliente é uma tarefa complexa e subjetiva, uma vez que mesmo entre humanos não existe consenso sobre qual estilo é mais ou menos apropriado para um determinado evento.

Portanto, a determinação de qual é o estilo de álbum fotográfico é o que mais se adequa aos propósitos do usuário transcende o escopo desta tese. Assim, um estilo de diagramação será escolhido para a abordagem a ser desenvolvida. Escolheu-se o BRIC,

o modelo proposto por Atkins (2008), devido aos seguintes motivos:

- não há restrições quanto ao número de fotografias por página;
- trata-se de um estilo cuja organização das fotos é retilínea, adequando-se a fotografias obtidas em eventos informais e formais;
- pode-se delimitar a região máxima da página a ser ocupada; e
- não há restrições quanto às dimensões das fotografias;

Apesar destas características positivas, há de se destacar também as desvantagens de utilização desta abordagem:

- não é possível, *a priori*, garantir a ordem das fotografias em uma mesma página;
- a região ocupada pelas imagens, precisa ser retangular;
- não é possível garantir que toda a área disponível será ocupada; e
- o resultado final depende da ordem de apresentação das fotografias;

Uma vez que as opções variam de acordo com o usuário e o método proposto é automático, não permitindo interações, foram definidas as seguintes restrições básicas no processo de formatação de um álbum fotográfico:

- fotografias cujos conteúdos estejam relacionados devem estar agrupadas;
- fotografias preferidas pelo usuário ou com qualidade estética superior devem ocupar maior área nas páginas do que as demais fotografias;
- o limite de páginas do álbum deve ser imposto pelo usuário e respeitado pelo sistema.

As restrições apresentadas estão de acordo com as restrições definidas por Atkins (2008). Em pesquisas mais recentes (e.g., Xiao et al. (2010), Chao et al. (2010), Atkins et al. (2010)), fundamentadas no estudo de Atkins (2008), verifica-se que o número de restrições consideradas praticamente se mantém.

Nesta pesquisa, é proposta uma expansão da abordagem de Atkins (2008), de modo a permitir que maior área disponível da página seja utilizada. Esta mudança, imperceptível em alguns casos, pode, entretanto, garantir ganhos significativos de espaço na área ocupada, a partir de recortes nas imagens que podem ser imperceptíveis. Tal mudança, contudo, contraria a restrição primária feita por Atkins (2008): a de preservar a proporção das fotografias. Para permitir que tal modificação seja possível, sem que o redimensionamento resulte na perda do aspecto original da imagem, propõe-se, neste capítulo, um algoritmo destinado ao recorte de fotografias.

Este capítulo está organizado conforme descrito a seguir. Na Seção 6.1, é descrita a extensão realizada sobre uma abordagem existente destinada à diagramação automática de álbuns fotográficos; Na Seção 6.2, apresenta-se um algoritmo para o recorte automático de fotografias com base no conteúdo das mesmas; A Seção 6.3 contém a abordagem utilizada para integrar ao BRIC o algoritmo destinado ao recorte automático de fotografias, com o intuito de aumentar a área útil das melhores fotografias no álbum. Por fim, na Seção 6.4, são feitas algumas considerações finais acerca do conteúdo deste capítulo.

6.1 Extensão da abordagem existente

Como pôde ser visto no Capítulo 2 (Fundamentação), diversas abordagens propostas para a diagramação automática de álbuns fotográficos têm sido documentadas na literatura da área, as quais, apesar de solucionarem, em diferentes graus, o problema, apresentam algumas limitações. Dentre tais limitações, podem-se destacar:

- A necessidade de intervenção do usuário;
- O pequeno número de modelos pré-estabelecidos;
- A distribuição do conteúdo na área útil;
- A inexistência de associação entre as fotografias;
- A resolução de problemas específicos;

- A inexistência de estratégias formais de validação (i.e., a partir da estruturação de experimentos e da realização de testes estatísticos) da qualidade visual do álbum fotográfico gerado.

Algumas destas restrições ainda não puderam ser completamente solucionadas (e.g., a divisão retilínea da área, uma vez que o formato típico das fotografias é retangular). Por outro lado, existem limitações que puderam ser solucionadas a partir de alterações nas restrições básicas de formatação do álbum.

A presente proposição inspira-se na pesquisa desenvolvida por Atkins (2008), cujos resultados objetiva-se incorporar as seguintes características:

1. Controle do dimensionamento de fotografias na página diagramada, de acordo com as preferências do usuário; e
2. Possibilidade de realização de recortes nas fotografias, a fim de permitir melhor visualização de seu conteúdo, especialmente naquelas que forem reduzidas.

Para que ambos os itens sejam satisfeitos, um dos requisitos primários definidos por Atkins (2008) teve que ser ignorado: “As proporções das imagens serão respeitadas”. A não conformidade a esta regra conferiu flexibilidade ao processo de posicionamento das imagens na página. Entretanto, a modificação da proporção de uma imagem poderia causar desconforto estético durante a visualização, uma vez sendo exibida fora de suas proporções originais.

Existem configurações de tamanhos de página e dimensões das imagens em que não há soluções que preencham toda a área disponível. Assim, será utilizado no sistema ora proposto o algoritmo destinado ao Recorte Automático de Fotografias descrito na Seção 6.2. Desta forma, pretende-se ampliar a área disponível da página com pequenas alterações na relação de aspecto (*aspect ratio*) da imagem.

Para evitar tal efeito, foi desenvolvido um algoritmo para o recorte automático de fotografias. Assim, podem-se modificar as dimensões de uma fotografia mantendo-se seus aspectos estéticos, assim como a qualidade da visualização, uma vez que interpolações são evitadas. O algoritmo de recorte leva em consideração o conteúdo da fotografia, de forma que seu alvo não seja acidentalmente recortado. O referido algoritmo é descrito na Seção 6.2.

6.2 Recorte automático de fotografias

Nesta seção, é descrito um método destinado ao recorte automático de fotografias. Com base no conteúdo da imagem, uma dada fotografia é recortada, até que se adeque a uma dada dimensão ou proporção.

6.2.1 Introdução

O problema do recorte de fotografias é comumente vinculado à adaptação da mídia a uma determinada proporção-alvo. A proporção de uma fotografia é a razão entre sua largura e altura. Tanto os dispositivos digitais de captura quanto aqueles destinados à exibição de imagens digitais tipicamente adotam uma proporção específica, a partir de um conjunto de possibilidades.

Entretanto, três proporções são mais comumente adotadas: (i) 4:3 (e.g., transmissão de televisão de definição padrão, monitores CRT); (ii) 3:2 (e.g., papel fotográfico); e (iii) 16:9 (e.g. tela *widescreen*, transmissão de televisão de alta definição). Para que uma mídia, em uma determinada proporção, possa ser renderizada em outra mídia, é preciso que haja algum processo de adaptação. Três métodos tradicionais para adaptação entre proporções diferentes são: tarjas (*pillars*), redimensionamento (*stretch*) e ampliação (*zoom*⁴) (OWEN; OWEN; HOLT, 2008; P., 2010).

O método das tarjas insere barras de cor sólida (denominadas de tarjas) do tamanho da diferença entre a menor das dimensões da imagem e as dimensões que a imagem deveria ter, caso estivesse na proporção-alvo. Por sua vez, o método do redimensionamento torna a imagem mais larga ou mais longa, de modo que a dimensão-alvo seja atingida. Finalmente, o método da ampliação adapta a imagem à dimensão-alvo, a partir de uma mudança de escala, seguida de um recorte centralizado na imagem.

Estes métodos de conversão frequentemente acarretam problemas na qualidade visual da imagem. Os dois primeiros métodos podem produzir imagens distorcidas ou

⁴O *zoom* pode significar tanto uma ampliação (*zoom in*) quanto uma redução (*zoom out*). Como nesta seção o método é aplicado para se adaptar uma imagem em um espaço de dimensões menores, sempre será dada, nesta seção, a conotação de *zoom in*. Por uma questão de simplicidade, no restante do capítulo será utilizado apenas o termo ampliação ou *zoom* para denotar *zoom in*.

com tarjas indesejadas. Por outro lado, o terceiro método preserva a proporção da imagem, embora o alvo da imagem possa ser cortado, caso nenhuma informação da imagem seja conhecida. Entretanto, tal método é o mais frequentemente adotado por empresas de revelação de fotografias e será denominado, neste documento, método automático de ampliação (*zoom*) e recorte (*automatic zoom & crop method - AZ&C*).

No contexto da fotografia digital, o principal objetivo dos métodos de recorte é adaptar automaticamente fotografias obtidas em uma determinada proporção a um espaço qualquer, em uma página de documento que, eventualmente, apresente uma proporção-alvo diferente.

Para a adaptação, será utilizado o método *zoom*, uma vez que este método não permite distorções na imagem e ocupa toda a área disponível sem a inserção de tarjas. Na medida do possível, a composição original da fotografia deve ser preservada⁵. Assim, pretende-se (1) maximizar a área preservada da fotografia e (2) evitar recorte do alvo. Apesar do método apresentado poder ser utilizado em qualquer imagem, melhores resultados são obtidos quando aplicados a retratos⁶.

Para evitar o recorte do alvo da fotografia, um conjunto de extratores de características é usado para identificar as regiões da imagem que possuem informações mais relevantes. Uma vez que não é possível saber, *a priori*, qual dos extratores de características provê informações mais confiáveis, um problema de otimização utilizando algoritmos genéticos foi formulado, visando à obtenção de um conjunto de pesos com os quais se podem combinar os extratores, de maneira que seja dada mais ênfase àqueles mais precisos.

Parte desta seção foi publicada na conferência SITIS (*International Conference on Signal-Image Technology & Internet-Based Systems*) 2010 sob o título *Combining Multiple Image Features to Guide Automatic Portrait Cropping for Rendering Different Aspect Ratios* (CAVALCANTI; GOMES; QUEIROZ, 2010a). Outra parte foi publicada no periódico *Journal of Multimedia Processing and Technology*, sob o título *An Exten-*

⁵Neste contexto, a composição de uma fotografia é a forma como o fotógrafo entendeu que deveriam se relacionar o alvo da fotografia e os demais elementos de menor importância.

⁶Neste documento, o termo *retrato* refere-se a fotografias nas quais o alvo principal são humanos.

ded Evaluation of Methods for Portrait Cropping. *Journal of Multimedia Processing and Technologies* (CAVALCANTI; GOMES; QUEIROZ, 2010b).

O método supradescrito será referenciado nesta pesquisa pela sigla MSCM (*Multiple Saliency Cropping Method*) ou método de recorte usando múltiplas saliências.

6.2.2 Método de recorte usando múltiplas saliências (*multiple saliency cropping method - MSCM*)

Nesta seção, será apresentada uma descrição geral do método de recorte usando múltiplas saliências (MSCM), constituído de dois estágios: (1) obtenção dos pesos para combinar os extratores de características, e (2) recorte da imagem.

Conforme discutido anteriormente, não foram encontradas, no âmbito desta pesquisa, outras abordagens para a adaptação de imagens com base no conteúdo (*Content-Aware Image Adaptation*) que (1) não necessitem a intervenção do usuário, (2) preservem a composição original da fotografia e (3) evitem distorções no alvo da fotografia. Portanto, o principal objetivo desta pesquisa foi a proposição de um algoritmo que, sem intervenção do usuário, recortasse uma fotografia, conferindo-lhe uma nova proporção, mas evitando distorções em seu alvo. Adicionalmente, objetivou-se não perder, na medida do possível, a composição original.

Dada uma imagem I de dimensões $l \times a$, a abordagem proposta consiste em obter uma imagem I' , cujas dimensões $l' \times a'$ serão estipuladas pelo usuário ou por um sistema automático de diagramação, sendo $l, a, l', a' \in \mathbb{N}_+$, $l' < l$ e $a' < a$.

A imagem I' é obtida após a remoção de $a - a'$ linhas e $l - l'$ colunas de I . As linhas e colunas a serem removidas serão aquelas que possuírem menor saliência. A saliência (representada por uma imagem $I_{saliencia}$) será determinada por extratores de características (cujas quantidade é n_{ec}), os quais serão combinados a partir da Equação 6.1.

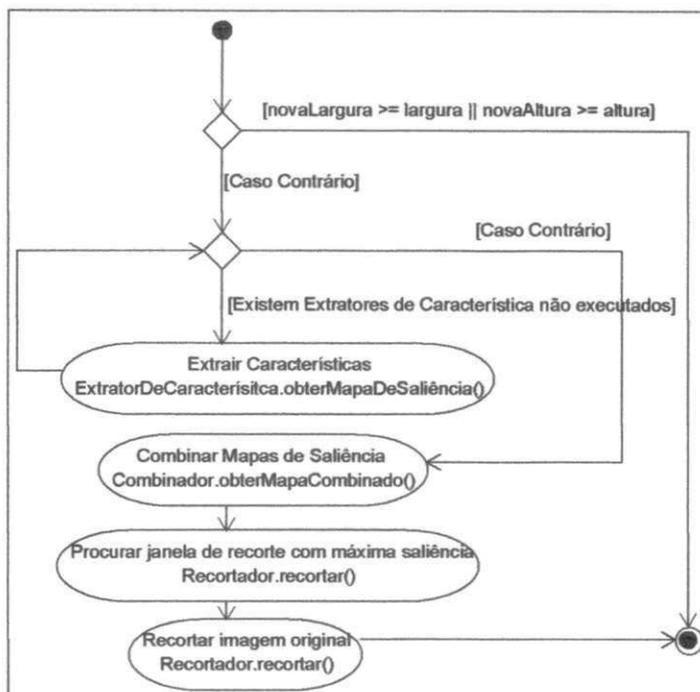
$$I_{saliencia} = \sum_{e=1}^{n_{ec}} EC_e(I) \times w_e, \quad (6.1)$$

em que $EC_e(I)$ é o resultado da execução do e -ésimo extrator de característica EC na

imagem I e w_e é o peso associado ao e -ésimo extrator.

Por se tratar do método mais importante do sistema, o método *recortar* encontra-se detalhado em um Diagrama de Atividades (ver Figura 6.1).

Figura 6.1: Diagrama de atividades representando o método *recortar* da classe *MSCM*.



O primeiro passo do processo é a geração de um mapa de saliência para cada extrator de características existente. Após sua geração, os mapas de saliência são combinados, a partir do método *obterMapaCombinado*, da classe *Combinador*. A combinação, conforme explicitada na Equação 6.1, requer pesos para cada característica. Esses pesos podem ser obtidos a partir de um treinamento usando algoritmos genéticos, o qual é realizado pela classe *GeradorDePesos* que, por sua vez, interage com a biblioteca de otimização, neste caso a biblioteca GALib (WALL, 2009), para a realização do treinamento.

Uma vez obtido o mapa de saliência combinado, o próximo passo é o recorte, efetuado pelo método *recortar*, da classe *Recortador*. Primeiramente, é localizada a região que contém o valor máximo da saliência e, em seguida, efetua-se o recorte da fotografia.

Nesta seção, o recorte de fotografias é apresentado como um estudo de caso de adaptação de mídia de uma proporção qualquer para uma proporção alvo. Todavia, o

algoritmo de recorte pode ser análogo se fossem informadas as dimensões-alvo ao invés de informada a proporção-alvo, pois o cálculo desta última nada mais é que o quociente entre a maior e a menor dimensões.

6.2.3 Extratores de características

O MSCM consiste em combinar informações obtidas de extratores de características e, em seguida, identificar as regiões de uma imagem que possuem maior número de elementos de interesse com base em tal combinação. Portanto, a principal parte desta abordagem depende dos extratores de características empregados. Quanto mais precisos forem os extratores de características na identificação do seu elemento de interesse, mais preciso tenderá a ser o recorte da fotografia. A estratégia de combinação de características geralmente traz benefícios de robustez, na medida em que eventuais falhas em um extrator podem ser compensadas por resultados satisfatórios de outros.

Os extratores de características descritos a seguir foram desenvolvidos com o objetivo de localizar: (1) alvos humanos, e (2) regiões relevantes da fotografia.

Localização de alvos humanos

A localização de alvos humanos em uma imagem não é uma tarefa trivial, dado o grande número de possíveis poses que o corpo ou face humanas podem assumir. Nesta pesquisa, duas abordagens foram utilizadas por apresentarem fortes evidências da presença e posicionamento de alvos humanos: detecção de tons-de-pele e detecção de faces.

Detecção de tons de pele (*skin-tone detector* - STD). Os tons da pele humana não são comumente encontrados na natureza. Portanto, um detector de tons de pele ou, simplesmente, detector de pele, pode ser útil na identificação de regiões da fotografia nas quais a probabilidade de existir pele humana é alta. Um detector de tons de pele é particularmente útil em situações nas quais outros detectores podem falhar, e.g., alguns detectores de face não são capazes de detectar faces capturadas de perfil.

Em abordagens comuns para STD são consideradas apenas, para fins de identificação, a informação de um único *pixel*, enquanto, em outras, é considerada, além de um *pixel* central, toda sua vizinhança. Vezhnevets, Sazonov e Andreeva (2003) realizaram um levantamento revisivo do estado da arte dos detectores de pele. O detector de

pele utilizado nesta pesquisa foi proposto por Cardoso (2009) e adota uma abordagem fundamentada somente na tonalidade do *pixel*. A abordagem utiliza redes neurais para a classificação dos *pixels* como pertencentes ou não a uma região de pele. Tal rede foi treinada a partir de imagens manualmente rotuladas. A informação transmitida como entrada da rede são os valores RGB do pixel, enquanto o valor de saída é um número binário.

Detecção de faces (*face detection* - FD). Faces são fortes evidências da presença humana e do seu posicionamento em uma fotografia. Atualmente, os detectores de faces apresentam baixas taxas de erro, porém ainda são incapazes de detectar, com grande precisão, certas poses, tais como faces em perfil, rotacionadas ou que apresentem oclusão significativa. Apesar de existirem limitações nos algoritmos de detecção de faces, vale a pena utilizá-los, uma vez que tais algoritmos são capazes de detectar rapidamente, de forma concisa e com precisão faces cuja pose seja frontal e vertical. Em tais casos, a presença de humanos é evidenciada pela detecção da face. Nesta pesquisa, foi utilizada uma implementação que tem por base o detector desenvolvido por Viola e Jones (2001). Melhorias nos detectores de face tendem a impactar positivamente esta pesquisa, uma vez que possibilitam a verificação da presença de alvos humanos com maior precisão.

Visando à obtenção de informações mais precisas quanto à existência de alvos humanos em uma imagem, razões antropométricas podem ser utilizadas para inferir as dimensões e o posicionamento de alvos dessa natureza, a partir das coordenadas da face. Tal abordagem foi verificada em pesquisas anteriores (CAVALCANTI, 2007; CAVALCANTI et al., 2006, 2010), tendo sido descrita na Seção 2.1.1.

Tendo em vista que as razões antropométricas podem apresentar falhas, no modelo utilizado nesta pesquisa as regiões indicadas por tal método são marcadas, empiricamente, com metade da certeza de uma face (e.g., se faces são encontradas, os *pixels* referentes são marcados com a intensidade 255, enquanto às regiões da imagem na qual são encontrados corpos é atribuída a intensidade 128). Tal escolha foi realizada empiricamente, uma vez que os observadores geralmente são intolerantes a recortes na face, enquanto é aceito um nível intermediário, definido nesta pesquisa em 50% apesar de outros valores serem possíveis, de recorte no corpo humano. Além disto, como

os detectores de face não disponibilizam as dimensões da cabeça, serão utilizadas as dimensões da face.

Localização de outras regiões relevantes

O desenvolvimento de um detector capaz de identificar qualquer tipo de objeto em uma cena permanece um desafio para pesquisadores da área de visão computacional. Melhores resultados têm sido obtidos quando detectores são desenvolvidos para uma classe particular de objetos (e.g., um detector de faces). Devido a tal dificuldade, detectores de informação de baixo-nível são frequentemente utilizados como uma solução aproximada. Nesta pesquisa, em particular, fez-se uso de um Detector de Bordas associado a um mecanismo de Atenção Visual.

Detector de bordas (*edge detection* - ED). Um detector de bordas (ED) pode ser utilizado com o intuito de identificar regiões de alta frequência em uma fotografia. Como regra geral de qualidade no campo da Fotografia, espera-se que o alvo encontre-se em uma região da fotografia que esteja focalizada. Regiões de alta frequência podem ser consideradas uma fraca evidência da presença de um objeto, pois regiões focalizadas apresentam bordas muito bem definidas.

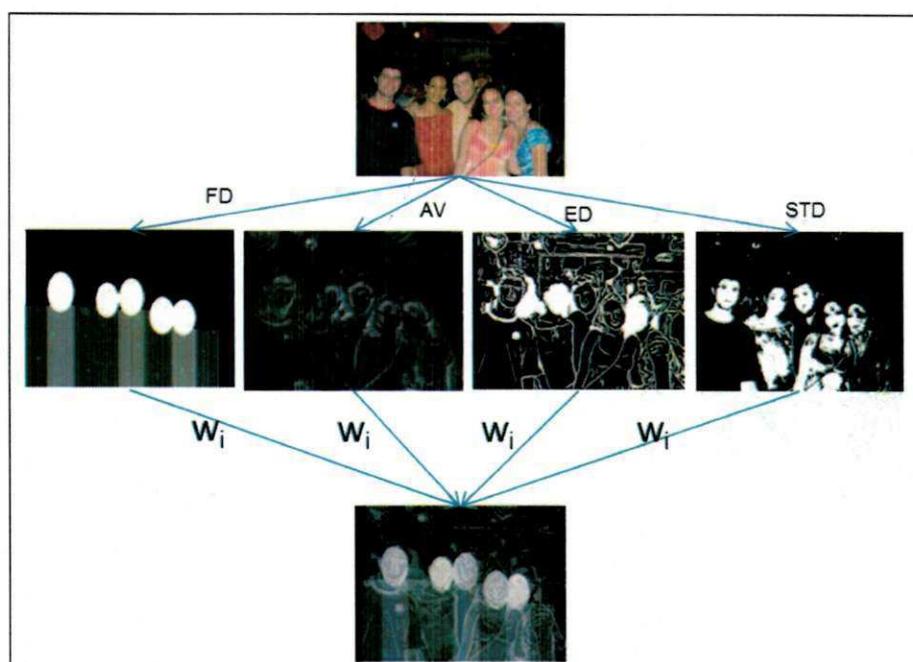
Foram testados diversos detectores de bordas, os quais não foram capazes de identificar corretamente as bordas do alvo. Para se identificar as bordas com menor fragmentação e melhor continuidade, um detector de bordas que utiliza a diferença de gaussianas a partir do fluxo (FDoG - *Flow-Based Difference of Gaussians*) (KANG; LEE; CHUI, 2007) foi incluído na abordagem. Com o objetivo de enfatizar as regiões correspondentes a objetos, os contornos fechados (delimitados pelas linhas encontradas via FDoG) serão preenchidos com uma cor sólida.

Atenção visual (*visual attention* - VA). Algoritmos de atenção visual (VA) inspiram-se no sistema visual humano, com o propósito de identificar a conspicuidade das regiões em uma cena. Considerando três planos de características (orientação, intensidade e cor), Itti, Rees e Tsotsos (2005) propuseram um algoritmo para combinar tais características e identificar as regiões de maior conspicuidade. O mecanismo de VA adotado no MSCM é inspirado na abordagem de Itti, Rees e Tsotsos (2005). A abordagem utilizada, entretanto, foi otimizada para a localização de alvos huma-

nos (PEREIRA; GOMES; FLORENTINO, 2006). Assim, o mecanismo de VA serve para evitar que regiões de importância, normalmente regiões salientes da imagem, sejam recortadas.

Na Figura 6.2, é mostrada a saída para cada um dos extratores de características supramencionados. Como entrada, todos os extratores de características admitem imagens na escala de cinza, com exceção do STD, que exige imagens coloridas. Todas as imagens geradas pelos extratores são monocromáticas.

Figura 6.2: Saídas para cada extrator de características: (a) a imagem original, (b) o FD com regiões antropométricas, (c) a VA (d) as bordas do ED, (e) STD, e (f) a soma das características.



Obtenção dos pesos das características

O conjunto de pesos associados à importância das características para a compreensão geral da fotografia deve ser ajustado, de modo a refletir o entendimento do problema por humanos. Tal estratégia é necessária, uma vez que áreas consideradas importantes por um extrator de características podem não corresponder àsquelas mais importantes de uma imagem sob o ponto de vista de um humano. Por exemplo, uma região que contém

uma quantidade elevada de bordas pode ter a mesma intensidade de outra que contém uma face, ao se observarem as saídas individuais dos extratores de características.

Assim, dada uma imagem de entrada, um conjunto de planos de características foi obtido. Este conjunto é composto por imagens nas quais os *pixels* com intensidade mínima (0) correspondem a regiões de nenhum interesse e *pixels* com intensidade máxima (255) correspondem a regiões de muito interesse, de acordo com a aquela característica. Podem existir, opcionalmente, até 254 outras gradações de interesse, de modo que o mapa de interesse consiste de uma imagem monocromática. Os extratores de características utilizados são aqueles supramencionados, a saber: FD, STD, VA e ED. A abordagem é modular e outros extratores podem ser facilmente adicionados ao sistema em extensões futuras.

O método da combinação ponderada de características utilizado é inspirado na abordagem de Pereira e Gomes (2006). Similarmente a tal abordagem, também são utilizados algoritmos genéticos (GA) para a obtenção de pesos destinados à localização de regiões de interesse. Diferentemente da abordagem supramencionada, não são consideradas unicamente as características de baixo nível.

Para a obtenção dos pesos que são utilizados na combinação de diferentes características, um conjunto de 100 imagens manualmente rotuladas foi preparado. Este número é compatível com o número utilizado em outras pesquisas documentadas na literatura que tratam do recorte automático de fotografias (ZHANG et al., 2005; SANTELLA et al., 2006).

As imagens foram rotuladas a partir da marcação de regiões retangulares consideradas mais importantes por dois observadores humanos⁷, i.e., as áreas que, preferencialmente, não deveriam ser recortadas durante a adaptação de uma imagem à nova proporção, de acordo com o observador. Cada retângulo corresponde à menor região que circunscribe o objeto por completo ou apenas parte de tal objeto, podendo ser utilizados vários retângulos para um alvo de maior complexidade.

⁷Externos à pesquisa.

Função de energia

Nesta seção, é definida a função de energia empregada no processo de otimização do GA. O principal objetivo era ajustar os pesos de cada característica, de modo que a soma dos planos ponderados de características resultasse em uma imagem cujas regiões mais salientes coincidisse com as áreas das imagens rotuladas por um humano como importantes (representados na imagem por *pixels* da cor branca).

Desta forma, sendo G uma imagem em escala de cinza de dimensões $m \times n$ e $G(x, y)$ o valor que corresponde à intensidade do *pixel* de G nas coordenadas (x, y) :

$$c(G) = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f(G(x, y)) \quad (6.2)$$

O processo de otimização consistiu na minimização dois tipos de erro: (1) o percentual de *pixels* rotulados por um humano como importantes, mas considerado como não importante pelo MSCM (Falso negativo ou erro tipo I, FN , explicitado na Equação 6.3); e (2) o percentual de *pixels* rotulados como importantes pelo MSCM que foram manualmente rotulados como não importantes (Falso positivo ou erro tipo II, FP , explicitado na Equação 6.4). Assim, as Equações 6.3 e 6.4 podem ser combinadas para a definição da função de energia Y ou acurácia, conforme a Equação 6.5.

$$FN = \frac{c(L - R)}{m \times n} \quad (6.3)$$

$$FP = \frac{c(R - L)}{m \times n} \quad (6.4)$$

$$Y = 1 - (FN + FP) \quad (6.5)$$

Tanto as imagens rotuladas por humanos quanto as imagens resultantes da abordagem proposta (L e R , respectivamente) tinham as mesmas dimensões, $m \times n$. A acurácia (Y) e os erros tipo I (FN) e II (FP) foram calculados com base na análise de concordâncias e discordâncias entre L e R .

Nas Equações 6.3 e 6.4, o sinal de subtração representa a diferença entre as intensi-

dades dos *pixels* da primeira para a segunda imagem. Por sua vez, a função c é definida na Equação 6.2 como sendo um somatório da função de limiarização f explicitada na Equação 6.6.

$$f(v) = \begin{cases} 1, & \text{se } v > 0 \\ 0, & \text{caso contrário.} \end{cases} \quad (6.6)$$

A imagem resultante (R) é calculada pela soma das imagens das características ponderadas pelos pesos previamente estimados para cada extrator de características, conforme explicitado na Equação 6.7, na qual k representa o número de características extraídas, I_t é a imagem resultante da execução do módulo t e w_t , o respectivo peso desta imagem na produção da imagem resultante R .

$$R = \frac{\sum_{t=1}^k w_t \times I_t}{\sum_{t=1}^k w_t} \quad (6.7)$$

A imagem rotulada (L) é obtida a partir da aplicação de um processo iterativo de rotulagem manual de coordenadas, representado por um conjunto \mathfrak{S} de pares ordenados, (x_1, y_1) e (x_2, y_2) . L é inicializada em 0 e, caso a soma das rotulagens resulte em valores maiores que 255, estes são limitados a 255. Deste modo:

$$L = L + d(L, M) \forall M \in \mathfrak{S}, \quad (6.8)$$

em que M contém os dois pares ordenados que definem um retângulo e d representa a função para colorir de branco o interior do retângulo indicado pelas coordenadas, conforme representado na Equação 6.9. Assim, sendo G uma imagem em escala de cinza:

$$d(G, M) = \begin{cases} G_{x,y} = 255, & \forall x, y | M_{x_1} < x < M_{x_2} \\ & \wedge M_{y_1} < y < M_{y_2}; \\ G_{x,y} = G_{x,y}, & \text{caso contrário.} \end{cases} \quad (6.9)$$

Após avaliar a acurácia do processo para cada imagem do conjunto de treinamento, a partir da Equação 6.5, a acurácia média \bar{Y} é obtida, sendo utilizada como critério de otimização.

Os pesos poderiam ser otimizados a partir de uma abordagem linear. Entretanto, para melhor explorar os conjuntos de possibilidades, uma abordagem utilizando algoritmos genéticos (GA) foi utilizada. Tal decisão tem por objetivo evitar a obtenção de um conjunto de pesos que corresponda a um máximo local, o qual pode ser encontrado utilizando métodos tradicionais de gradiente. O paralelismo intrínseco encontrado nos GA pode evitar o máximo local, graças ao processo pseudo-paralelo de busca de soluções sub-ótimas.

Outras abordagens poderiam ser utilizadas, entretanto, uma vez que o foco desta pesquisa não é a investigação aprofundada de métodos de otimização, tal investigação fica como sugestão para investigações futuras e.g., otimização por enxames de partículas (*Particle Swarm Optimization* - PSO). Uma comparação mais aprofundada sobre os motivos de utilização dos GA ao invés do PSO encontra-se na Seção A.2.

Algoritmos genéticos (*genetic algorithms* - GA). Algoritmos genéticos foram utilizados para a obtenção de um conjunto de pesos que tornassem a rotulagem automática o mais próximo possível da rotulagem humana, no tocante às regiões relevantes da imagem. A decisão da utilização de algoritmos genéticos é comentada na Seção A.2. Cada genoma representa um conjunto de pesos, o que significa que quanto mais a imagem resultante coincidissem com a marcação humana, mais apto seria o indivíduo possuidor de tal genoma. A meta da função objetivo utilizada pelo algoritmo genético era maximizar a acurácia (ou, equivalentemente, minimizar a soma dos erros FP e FN indicados acima).

A ferramenta GALib (WALL, 2009) foi utilizada para dar suporte ao processo de otimização pelo GA. Existiam $k + 1$ fenótipos dos quais o primeiro correspondia ao limiar de binarização da imagem e os demais k correspondiam aos pesos atribuídos a cada imagem resultante de um extrator de características. A população, em estado estacionário, consistiu em 300 indivíduos e evoluiu por 200 gerações, com uma probabilidade de mutação de 5% e probabilidade de cruzamento de 70%. A recombinação dos indivíduos utilizada foi a de um ponto, enquanto a seleção foi realizada pelo método da roleta. O critério de parada adotado foi o número de gerações ou a convergência média de 85%.

Outras configurações foram testadas, a partir de uma variação sistemática destes pa-

râmetros, mas a configuração supraexplicitada forneceu os melhores resultados, dentre as configurações testadas. Os testes realizados variaram a probabilidade de cruzamento entre 0,10 e 1, com incrementos de 0,10 e a probabilidade de mutação entre 1% e 21%, com incrementos de 5%. Adicionalmente, variou-se o número de gerações entre 5 e 500, com passos de 20 e a população entre 20 e 1100, com passos de 40. Com este melhor conjunto de parâmetros, o teste foi repetido por 10 vezes, resultando em resultados similares. Após 200 gerações, a melhor combinação de pesos resultou em uma acurácia de 73,89%.

A otimização a partir de GA é um experimento caro, no tocante ao tempo de execução do processo. Entretanto, GA são utilizados apenas no processo de obtenção de pesos, não sendo necessário a re-execução do processo de aprendizagem uma vez concluída a otimização.

Os pesos obtidos a partir dos extratores de características utilizados são apresentados na Tabela 6.1.

Tabela 6.1: Pesos obtidos pelos extratores de características utilizados.

FD	VA	ED	STD
0,73	0,46	0,33	0,13

Pode-se perceber, a partir de tais resultados, que o detector de faces teve o maior peso na produção do mapa de saliência. Tal resultado era esperado uma vez que esta pesquisa é focada em fotografias contendo alvos humanos.

Recorte da imagem final

Após a obtenção do conjunto de pesos, o próximo passo consistiu do recorte da imagem final, cujo algoritmo é descrito nesta seção (conforme Equação 6.7). Uma vez que o algoritmo para o recorte de imagens é similar em ambas as orientações, por questões de simplicidade, mas sem perda de generalidade, descreve-se aqui apenas a estratégia adotada para a orientação horizontal. Os cálculos para recortes da orientação vertical são análogos, sendo necessária apenas a substituição das orientações dos termos empregados no cálculo (e.g., substituição das colunas pelas linhas).

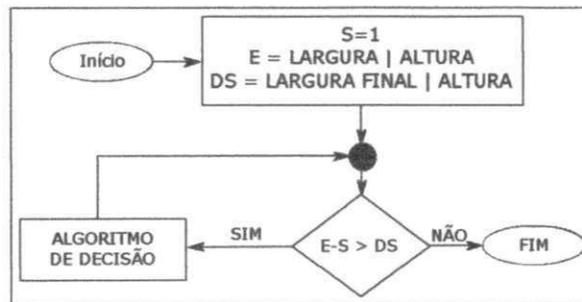
A principal restrição desta abordagem reside em determinar a nova proporção com o mínimo de modificações da imagem original. Com o intuito de tornar tal restrição possível, a maior dimensão é mantida (no caso da orientação horizontal, a largura será sempre a maior dimensão). A partir de tal restrição, a decisão resume-se a qual coluna da imagem descartar em cada passo do algoritmo. O cálculo da nova altura da imagem (h) é realizado conforme a Equação 6.10.

$$h = W \times a, \quad (6.10)$$

em que a representa a proporção desejada e W representa a largura original da imagem, e.g., para se obter uma imagem com proporção 16:9, $a = 16 \div 9 \simeq 1,7777$.

O algoritmo destinado à seleção de qual coluna deverá ser recortada (da esquerda ou da direita) a cada passo do algoritmo é descrito a seguir. Um fluxograma resumido do algoritmo de recorte é ilustrado na Figura 6.3, na qual S e E denotam, respectivamente, as colunas de início e fim da nova imagem. O algoritmo de decisão determina a coluna a ser descartada.

Figura 6.3: Síntese gráfica do algoritmo de recorte.



Foram investigadas quatro abordagens destinadas à tomada de decisão sobre qual coluna descartar. Na primeira, levou-se em consideração a intensidade máxima, comparando-se as colunas adjacentes às bordas, sendo descartada a coluna com menor intensidade. Se as intensidades forem iguais, a próxima coluna é analisada, repetindo-se tal etapa até que uma coluna com menor intensidade seja encontrada. Tal abordagem visa a preservar o número de colunas com maiores intensidades, as quais representam regiões consideradas importantes pelos extratores de características.

Na segunda abordagem, leva-se em consideração a soma dos *pixels* de uma coluna. Tal abordagem favorece colunas com um número maior de regiões importantes, ao invés de considerar apenas aquelas que contêm pelo menos um *pixel* com nível de cinza elevado. As colunas adjacentes às bordas são comparadas, similarmente ao que ocorre na abordagem anterior. Entretanto, erros dos extratores podem comprometer sensivelmente a qualidade do recorte.

Na terceira abordagem, as duas abordagens anteriores são combinadas: primeiro, somam-se as intensidades dos *pixels* e, em seguida, compara-se o valor máximo com um limiar pré-determinado. Entretanto, pôde-se perceber que, ao considerar uma única coluna, era possível que uma pequena região contivesse alta intensidade, forçando o recorte na direção oposta. Para evitar resultados indesejados, propôs-se levar em consideração não apenas uma coluna, mas uma janela de análise. Assim, ao invés de se comparar as somas em uma única coluna, a soma de várias colunas foi considerada.

A quarta e última abordagem difere das anteriores por considerar a composição inicial da fotografia. O objetivo é posicionar o alvo da fotografia na imagem final de forma semelhante ao posicionamento na imagem original.

O primeiro passo consistiu em encontrar áreas de interesse na imagem original. Para acelerar o cálculo, imagens integrais (VIOLA; JONES, 2001) foram utilizadas para calcular a saliência acumulada em áreas candidatas. Uma vez que o sistema de recorte proposto preserva uma das dimensões da imagem, é necessário apenas obter uma matriz de somas, e.g., A , com dimensão $1 \times w$. Como a dimensão candidata f será menor que a dimensão de A , serão procuradas, então, as 10% maiores somas dentre todos os possíveis candidatos em A cujas dimensões sejam f .

O segundo passo consistiu em encontrar qual dos candidatos se assemelhava mais à imagem original, de acordo com a distribuição espacial da saliência. Para selecionar a área candidata final, realizou-se o cálculo expresso pela Equação 6.11.

$$m = \arg \min_{m_i} f(m_i, m^*), m \in H'(w), \quad (6.11)$$

em que $H'(w)$ é um conjunto formado por todos os subconjuntos de H , com largura w (sendo H a função para cálculo do histograma de projeção de uma imagem descrita na

Equação 4.1) e f uma função para o cálculo da diferença absoluta entre os quartis de dois histogramas de projeção médios.

Durante a realização dos treinamentos, pôde-se avaliar a qualidade das quatro abordagens supramencionadas. Devido aos melhores resultados e à melhor eficiência computacional, o quarto método foi empregado nesta pesquisa. Uma investigação sobre a influência dos métodos no julgamento das imagens resultantes por observadores humanos também pode ser apontado como tema de investigação futura.

Considerações da seção

Nesta seção, foi apresentado um método destinado ao recorte automático de fotografias com respeito ao conteúdo: o MSCM. Este método (CAVALCANTI; GOMES; QUEIROZ, 2010a, 2010b) foi concebido com o intuito de adaptar fotografias obtidas por consumidores para diferentes proporções sem a necessidade de intervenção. Assim, o objetivo desta pesquisa é mostrar que a combinação de múltiplas características é essencial ao processo de recorte, uma vez que as características, isoladamente, podem não ser suficientes para a realização de um recorte adequado. Isto pôde ser verificado por meio de uma avaliação experimental, na qual o resultado do processo de recorte pelo MSCM foi comparado a resultados produzidos por métodos que se utilizam das características separadamente.

Foi concluído que, em 80% dos julgamentos, os quais foram realizados em 100 imagens por 12 participantes sem conhecimento dos propósitos do experimento, o MSCM foi considerado melhor do que ou similar aos demais métodos. Também foi possível inferir que, a partir do conjunto de imagens considerado, os resultados produzidos a partir da característica de Detecção de Faces (FD), isoladamente, são equivalentes ao método AZ&C.

Uma das desvantagens do MSCM reside no processamento daquelas imagens cuja composição fotográfica não foi realizada adequadamente pelo fotógrafo, o que resulta em permitir que o tema da fotografia esteja demasiadamente próximo de ambas as bordas e nenhum conteúdo importante no meio da imagem. Neste cenário, algoritmos de redirecionamento devem, provavelmente, produzir melhores resultados.

Uma vez que o MSCM mostrou-se adequado para recortes de fotografias, este pode

ser utilizado como ferramenta auxiliar para a diagramação de álbuns fotográficos, o qual será discutido na Seção 6.3.

6.3 Integração do sistema de recorte ao diagramador automático

Conforme mencionado anteriormente, alterações na relação de aspecto *aspect ratio* das imagens a serem diagramadas podem permitir que maior área da página seja ocupada pelas fotografias. Uma vez que os recortes são orientados ao conteúdo, conforme descrito na seção anterior, a tendência é que as áreas de interesse das imagens tornem-se mais evidentes à medida que maior área da página é ocupada.

Como o algoritmo BRIC é iterativo, não é possível, *a priori*, saber qual o nível de recorte necessário para a obtenção de uma árvore que ocupe a maior área possível. Como trata-se de um problema de obtenção de um máximo, decidiu-se por utilizar um algoritmo de otimização. Neste caso, optou-se pela utilização dos algoritmos genéticos, pelos motivos mencionados anteriormente.

Dessa forma, cada indivíduo da evolução consistirá no percentual de recorte horizontal e vertical que uma imagem pode ter. Assim, cada imagem possui 2 valores, totalizando $k \times 2$ valores, sendo k o número de imagens existentes na página. Por sua vez, a função custo é traduzida pela área ocupada pela árvore contendo as imagens com as novas proporções. Para os testes, foram evoluídos 40 indivíduos por 10 gerações, com substituição de 50% da população por geração e probabilidade de cruzamento de 50%. Tais valores foram adotados por se pretender cobrir um número de possibilidades do espaço de busca sem que fossem necessários muitos testes, já que se tinha como objetivo a produção de uma ferramenta a ser utilizada rapidamente, conforme discutido anteriormente, o que não seria possível caso se optasse por um maior número de testes.

Vale salientar que o nível de recorte para cada imagem possui um valor máximo. Para os experimentos realizados, foi definido que o máximo de recorte permitido seria de 10% da largura e da altura. Logo, o valor do indivíduo, na evolução genética, equivale ao percentual dos 10% de recorte que será aplicado à imagem. Outra abordagem testada consistiu em definir o percentual máximo de recorte, de acordo com o possível interesse

do usuário pela fotografia, o qual pode ser avaliado por funções como aquelas definidas no Capítulo 4. Entretanto, para o experimento descrito na Seção 7.4, foi utilizada a primeira abordagem, uma vez que tal abordagem garante recortes, na maioria dos casos, imperceptíveis, mas que dão flexibilidade ao algoritmo diagramador.

Como o algoritmo do BRIC é iterativo e descarta as árvores que produziram pior resultado à medida que novas imagens são inseridas no cálculo, algumas configurações que seriam melhores para um número maior de imagens poderão ser descartadas, uma vez que para valores mais baixos apresentarão resultados inferiores, já que apenas as L melhores árvores são utilizadas na iteração seguinte. Para garantir que uma árvore potencialmente promissora não seja descartada em etapas preliminares da iteração, decidiu-se alterar o valor de L de 4 (conforme utilizado na abordagem de Atkins (2008)) para 10.

Para aumentar a possibilidade de obtenção de árvores com resultados satisfatórios, decidiu-se aumentar o valor de L à medida que melhores resultados não fossem obtidos em iterações posteriores. Por exemplo, caso não fossem encontradas árvores que resolvessem o problema em uma dada iteração, retornava-se à primeira iteração com um valor maior de L . Um valor de L maior que 10 não foi escolhido como valor padrão porque aumentaria a complexidade do problema, dado que, apesar de apenas L árvores serem utilizadas na iteração seguinte, um grande número de árvores poderia ser criado para cada uma das L árvores, dependendo do número de folhas que cada árvore possuísse. Conforme a alteração, o valor de L só aumentará se não forem encontradas árvores que consigam resolver o problema.

Em virtude do cálculo da função de custo estar relacionada à área total ocupada da página, alguns resultados insatisfatórios poderiam ser obtidos, e.g., uma imagem ocupando grande área disponível na página e as demais imagens com dimensões muito inferiores. Para evitar esse tipo de comportamento, foi necessário adicionar uma regra empírica ao algoritmo tradicional do BRIC: a diferença da proporção entre a menor e a maior imagem em uma mesma página não pode ser menor que 1:5.

6.4 Considerações finais

Neste capítulo, foi apresentado um método para a Diagramação Automática de Álbuns Fotográficos. Inspirado na pesquisa de Atkins (2008), o qual posiciona fotografias em um espaço indicado, de maneira que aquelas rotuladas pelo usuário como as mais importantes ou de maior qualidade estética ocupem maior área no espaço considerado.

Para tanto, um algoritmo de recorte automático de fotografias foi desenvolvido e integrado ao sistema. Tal algoritmo permite que as fotografias sejam adaptadas a um espaço disponível na página sem perda de qualidade, uma vez que o recorte é orientado ao conteúdo.

Os próximos passos são: (1) o desenvolvimento de algoritmos que gerem automaticamente álbuns fotográficos respeitando as restrições ainda não resolvidas (e.g. o agrupamento de fotografias temáticas ou cronologicamente afins); e (2) o desenvolvimento de um algoritmo destinado à geração automática de álbuns fotográficos, utilizando aprendizagem de máquina. No segundo passo, a conformidade às regras e a qualidade estética podem constituir a função objetivo e cada combinação de distribuições e dimensões das fotografias no espaço das páginas são as soluções a serem testadas.

Ao final do desenvolvimento do sistema, este deverá ser verificado por meio de avaliações subjetivas e testes estatísticos.

Capítulo 7

Experimentos

Neste capítulo, são descritos os experimentos realizados para a validação dos métodos propostos nesta tese. Este capítulo divide-se em seções, cada uma das quais contendo a descrição de uma etapa da pesquisa. Na Seção 7.1, é descrita a avaliação experimental destinada à validação do método para o recorte automático de fotografias. Em seguida, na Seção 7.2, são apresentados os experimentos realizados para avaliar o algoritmo de obtenção de segmentadores por cor a partir da comparação do resultado da extração das cores com o processo realizado por humanos. Em seguida, na Seção 7.3, é descrito o método utilizado para avaliar a qualidade do algoritmo identificador de fotografias similares ou duplicatas. Por fim, na Seção 7.4, é apresentado o experimento destinado à validação do sistema completo, o qual envolve a seleção de fotografias e a diagramação de álbuns fotográficos.

7.1 Avaliação experimental do método de recorte automático de fotografias

Esta seção divide-se em duas partes: na primeira parte (Seção 7.1.1), são apresentados os experimentos preliminares (em que se comparam apenas as versões canônicas dos métodos estudados), enquanto na segunda parte (Seção 7.1.2) experimentos e resultados mais detalhados são apresentados (em que há uma exploração mais ampla dos métodos

utilizados).

7.1.1 Experimentos preliminares

Com o objetivo de analisar o MSCM, uma avaliação subjetiva preliminar foi realizada. Uma vez que não foi encontrada, durante a revisão bibliográfica feita no âmbito desta pesquisa, uma estratégia objetiva para a avaliação qualitativa de recortes em fotografias, tal avaliação foi realizada a partir de uma comparação subjetiva entre o MSCM e as abordagens previamente descritas, as quais são comumente utilizadas em processos de recorte de fotografias, i.e., a abordagem de Suh et al. (2003), o método Seam-Carving (AVIDAN; SHAMIR, 2007) e o método AZ&C.

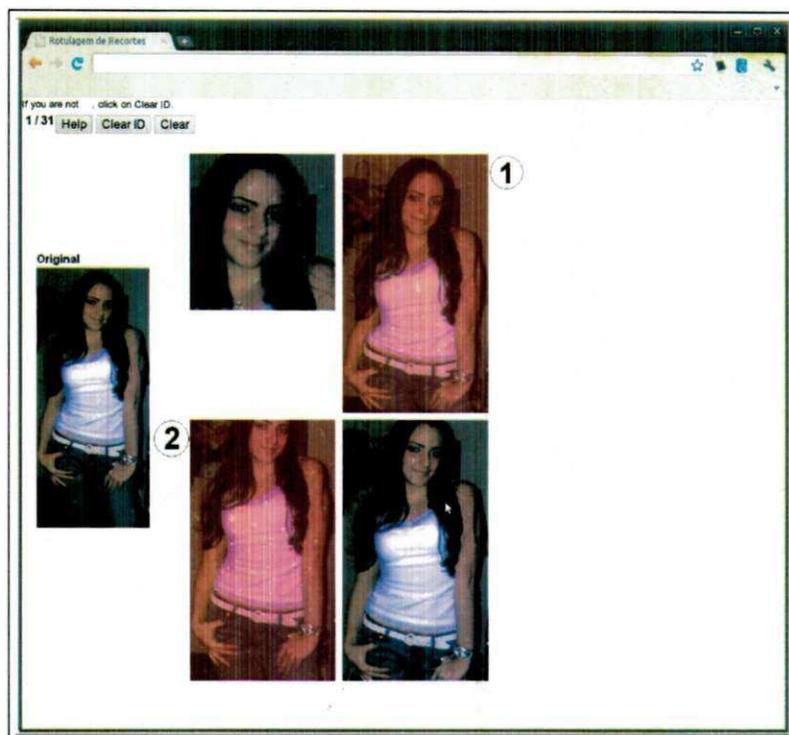
Trinta fotografias contendo alvos humanos (todas diferentes do conjunto de imagens utilizado no processo de otimização do GA) foram recortadas por intermédio dos quatro métodos avaliados, de modo que as imagens se adequassem às proporções 16:9, 3:2 e 4:3. Contudo, na abordagem de Suh et al. (2003), não é possível definir a proporção-alvo, sendo a saliência o único fator considerado para o recorte. Desta forma, em alguns casos, pode haver, por parte deste método, um recorte maior, em comparação aos demais métodos. O número de imagens foi definido com base nos requisitos de avaliação subjetiva do experimento, dada a dificuldade de recrutamento de participantes voluntários representativos, caso a quantidade de fotografias fosse significativamente maior. As fotografias foram cuidadosamente escolhidas, de maneira a representar situações típicas, tanto quanto ao número de alvos humanos, como quanto às posições e poses de tais alvos nas fotografias.

O número de participantes também foi limitado devido à necessidade de tempo requerido pelos participantes para conseguirem observar minuciosamente as fotografias. Um total de 35 voluntários participaram do experimento.

As fotografias foram redimensionadas, cada uma a partir de um dos quatro métodos, posicionadas lado-a-lado e apresentadas em um formulário *Web*. No total, 5 imagens foram exibidas (a imagem original e as quatro imagens redimensionadas a partir de cada método) para os participantes, aos quais solicitou-se a ordenação dos recortes

do melhor para o pior. Para assegurar a imparcialidade da rotulagem, a ordem de apresentação das imagens resultantes para cada proporção e para o acesso de cada participante foi randomizada. O formulário Web utilizado é exibido na Figura 7.1.

Figura 7.1: Tela de captura da página Web usada para obter a opinião dos participantes sobre a qualidade dos recortes realizados, ordenando-os do melhor (1) para o pior (4).



Ao invés de analisar os resultados a partir da verificação de qual dos métodos foi classificado como o melhor em um número maior de vezes, uma abordagem diferente foi adotada pois, em muitos casos, diferentes métodos geraram resultados quase indistinguíveis, não permitindo uma escolha consistente por parte dos participantes. Para minimizar os efeitos deste tipo de cenário, cada método recebeu um peso, de acordo com a ordem de seleção: os escores atribuídos pelos participantes variaram, de forma decrescente de acordo com a preferência do usuário, entre 4 e 1. Para cada método, o escore final s foi calculado a partir da soma do produto entre o número de participantes e o escore que atribuíram a cada imagem, assim como definido na Equação 7.1.

$$s = \sum_{i=1}^n p(i) \times r, \quad (7.1)$$

em que n representa o número de imagens e p corresponde ao número de participantes que atribuíram a uma dada imagem i o escore r . O percentual que indica a preferência dos observadores foi, então, obtido a partir da divisão entre o escore final s de cada método e a soma dos escores de todos os métodos. Na Tabela 7.1, são mostradas as preferências dos participantes neste experimento.

Tabela 7.1: Resumo do desempenho do método.

Método	Escore Final
AZ&C	3333 (30,98%)
MSCM	3012 (28,00%)
Suh et al. (2003)	2221 (20,65%)
Seam-Carving (AVIDAN; SHAMIR, 2007)	2191 (20,37%)

O Teste ANOVA fator único (WALPOLE et al., 2008) foi administrado, para verificar a existência de diferenças entre as preferências dos votantes com relação às abordagens consideradas. Em complementação ao teste ANOVA, foi administrado o teste de Tukey-Kramer (WALPOLE et al., 2008), para executar comparações pareadas. A administração de tal teste permitiu identificar para quais pares de métodos as diferenças foram significativas, conforme se pode verificar na Tabela 7.2.

Tabela 7.2: Resultados do teste de Tukey-Kramer.

Comparação Pareada	Diferenças Significativas
MSCM vs AZ&C	NÃO
MSCM vs Suh et al.	SIM
MSCM vs Seam-Carving	SIM
Suh et al. vs Seam-Carving	NÃO
Suh et al. vs AZ&C	SIM
AZ&C vs Seam-Carving	SIM

Os resultados do teste de Tukey-Kramer apontaram a existência de diferenças significativas tanto entre o MSCM e o método de Suh et al. (2003) quanto entre o MSCM

e o método *Seam-Carving* (AVIDAN; SHAMIR, 2007). Também se verificou que não existiam diferenças significativas entre os métodos MSCM e o AZ&C. Portanto, existe evidência suficiente para rejeitar a hipótese nula (os métodos são equivalentes)

Os resultados obtidos eram esperados, por algumas razões. Conforme previamente discutido, o método *Seam-Carving* (SC) (AVIDAN; SHAMIR, 2007) produz imagens distorcidas em algumas circunstâncias, e.g., quando o alvo ocupa um grande espaço na fotografia. Entretanto, o método SC não é significativamente superior nos demais cenários. A abordagem de Suh et al. (2003) destinou-se, originalmente, ao recorte de miniaturas. Logo este método desconsidera usualmente o conteúdo da fotografia, preservando apenas a informação essencial - faces ou grandes áreas salientes, o que explica porque, mesmo em um cenário no qual apenas existem faces, a abordagem produz resultados piores. Finalmente, o método AZ&C produz melhores resultados, especialmente devido à tendência de centralização do alvo por parte dos fotógrafos amadores. Portanto, no cenário considerado, os métodos MSCM e o AZ&C produziram recortes similares.

Uma vez que a análise dos dados mostrou que não existem diferenças estatisticamente significativas entre os métodos MSCM e o AZ&C, um novo experimento foi projetado, com o propósito de comparar estes dois métodos isoladamente, considerando tais métodos em uma base de fotografias com aparência mais profissional, nas quais os alvos tendem a não estar centralizados ou interagem com outros elementos da fotografia.

Para evidenciar as diferenças entre ambos os métodos, um conjunto mais amplo de fotografias foi estruturado, totalizando duzentas fotografias. Este elevado número de fotografias, contudo, comprometeu a participação de um número amplo de participantes voluntários, de modo que se contou apenas com a participação de 7 voluntários, os quais foram instruídos a marcar a opção que refletia suas preferências, dentre as seguintes opções: (1) a imagem da esquerda é muito melhor do que a imagem da direita; (2) a imagem da esquerda é ligeiramente melhor do que a imagem da direita; (3) as imagens parecem iguais; (4) a imagem da direita é ligeiramente melhor do que a da esquerda;

(5) a imagem da direita é muito melhor do que a imagem da esquerda. Em seguida, as imagens foram exibidas nas proporções 4:3, 3:2 e 16:9, resultantes da execução dos métodos MSCM e AZ&C.

Definir o que é um bom recorte é uma tarefa subjetiva. Entretanto, pode ser simples identificar um mau recorte, o qual frequentemente ocorre quando o alvo da fotografia é cortado em áreas importantes, tais como a cabeça ou as articulações. É mais difícil identificar qual o melhor dentre dois recortes quando, em nenhum deles, o alvo não está cortado, ainda que ambos os recortes sejam substancialmente diferentes. A grande dificuldade é a realização de comparações quando os recortes são muito próximos. Eventualmente, um dos recortes pode ter aparência global superior, sendo, contudo, seu concorrente igualmente apropriado.

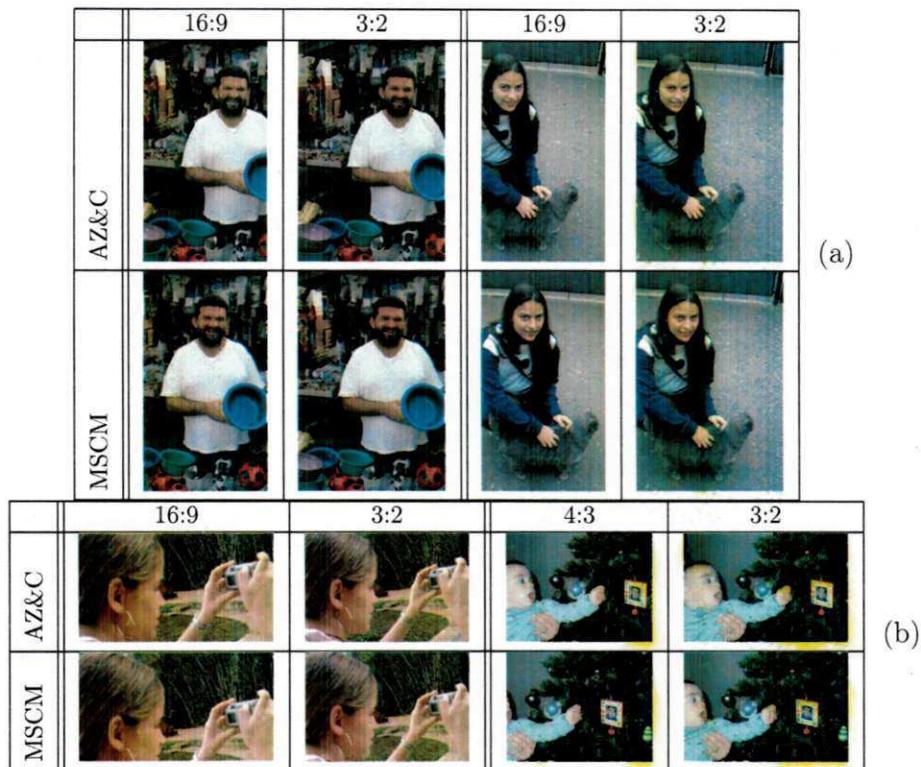
Um escore entre -2 e 2 foi atribuído a cada comparação, sendo os valores positivos correspondentes a votos favorecendo imagens produzidas pelo método MSCM e os valores negativos correspondentes a votos favorecendo o método AZ&C. A gradação de escores deveu-se aos motivos anteriormente apresentados, ou seja, pela dificuldade de se julgar um recorte isoladamente. A partir da gradação, o votante pode demonstrar que gostou mais de um recorte mas que considerou o outro apropriado.

A avaliação subjetiva mostrou que, em 77% das fotografias (154 fotografias), os participantes consideraram a fotografia recortada pelo MSCM melhor do que a fotografia produzida pelo método AZ&C ou similar a ela. Quando considerando apenas fotografias ranqueadas como melhores do que aquelas produzidas pelo método AZ&C, o percentual obtido foi de 60% (120 fotografias). Em apenas 23% das fotografias (46 fotografias) o método AZ&C foi considerado superior.

Estes resultados mostram que o método AZ&C não produz erros grosseiros, conforme poderia se esperar. Dado que grande parte dos fotógrafos amadores utiliza a regra da centralização do alvo, um recorte central costuma ser uma abordagem eficaz. O MSCM produz excelentes resultados para as demais fotografias, nas quais o alvo encontra-se ligeiramente desviado do centro, resultado não obtido pelo método AZ&C, nem pelas demais abordagens testadas.

Dois exemplos de fotografias rotuladas como “muito melhores” do que o método AZ&C e um exemplo no qual os recortes foram considerados piores do que o método AZ&C são ilustrados na Figura 7.2. São mostrados exemplos de adaptação para duas proporções alvo: 16:9 e 3:2.

Figura 7.2: Exemplo da saída dos métodos MSCM e do AZ&C: em (a) o MSCM foi considerado melhor do que o AZ&C, em (b) o MSCM foi considerado pior pelos participantes.



7.1.2 Avaliação estendida

Um último experimento foi então elaborado e conduzido, com o objetivo de verificar a importância da combinação das características existentes. Este experimento foi similar àquele descrito na Seção 7.1.1, uma vez que considerou apenas os métodos mais bem ranqueados: o AZ&C e o MSCM. Adicionalmente a estes dois métodos, quatro outros métodos foram considerados. Apesar de se fundamentarem na mesma abordagem de recorte, com base na saliência descrita na Seção 6.2.3, na qual foram utilizados os

mesmos mapas de saliência combinados, em cada método foi utilizada apenas uma das quatro características (Detecção de Faces, Atenção Visual, Detecção de Tom-de-Pele e Detecção de Bordas). Assim, foram criados quatro métodos, referentes a cada uma das características: FD', VA', STD' e ED'.

Com o objetivo de utilizar as características individualmente, o método de recorte sofreu alguns ajustes. Por exemplo, quando nenhuma face era detectada e o mapa de saliência das faces era usado, o método de recorte era programado para se comportar similarmente ao método AZ&C.

Para todas as imagens, a proporção-alvo foi 16:9, uma vez que tal proporção forçava a existência de maiores recortes em todos os métodos, para as fotografias utilizadas no experimento (originalmente, em sua maioria, na proporção 4:3). A opção por maiores recortes deveu-se ao fato de que, para pequenos recortes, as fotografias resultantes apresentavam um grau de similaridade tão elevado que comprometia o julgamento do participante.

Um total de 100 imagens foi recortado por seis diferentes métodos. O conjunto de fotografias foi estruturado com o intuito de ser heterogêneo, de modo a possibilitar a avaliação da robustez do algoritmo e, ao mesmo tempo, refletir o teor de fotografias comumente utilizadas, por exemplo, pela ampla maioria de fotografias contendo alvos humanos.

O conjunto conteve fotografias com e sem indivíduos como alvos. As fotografias cujo alvo principal eram humanos continham pelo menos um e, no máximo, três alvos humanos. Nas demais fotografias, o alvo principal era constituído por objetos (e.g., aviões, estátuas), animais ou paisagens.

O conjunto também foi estruturado de modo a variar quanto ao nível de *zoom* aplicado, contendo aproximadamente o mesmo número de fotografias em *close-up* (C-Up), Média Distância (MD), Longa Distância (LD) e só uma pequena quantidade de fotografias em *Close-Up* Extremo (EC-Up). Aproximadamente metade das imagens do conjunto continha alvos humanos em pose frontal e, na outra metade, estes encontravam-se em outra pose.

A distribuição dos conjuntos é mostrada na Tabela 7.3.

Tabela 7.3: Distribuição dos tipos de fotografia usados no experimento estendido.

Tipo	Distribuição
Alvo principal	Humano: 94%, Não-humano: 6%
Pose do alvo	Frontal: 43%, Não-frontal: 57%
Nível de <i>zoom</i>	EC-Up: 5%, C-Up: 36%, MD: 37%, LD: 22%
Número de indivíduos	1: 67%, 2: 24%, 3+: 3%

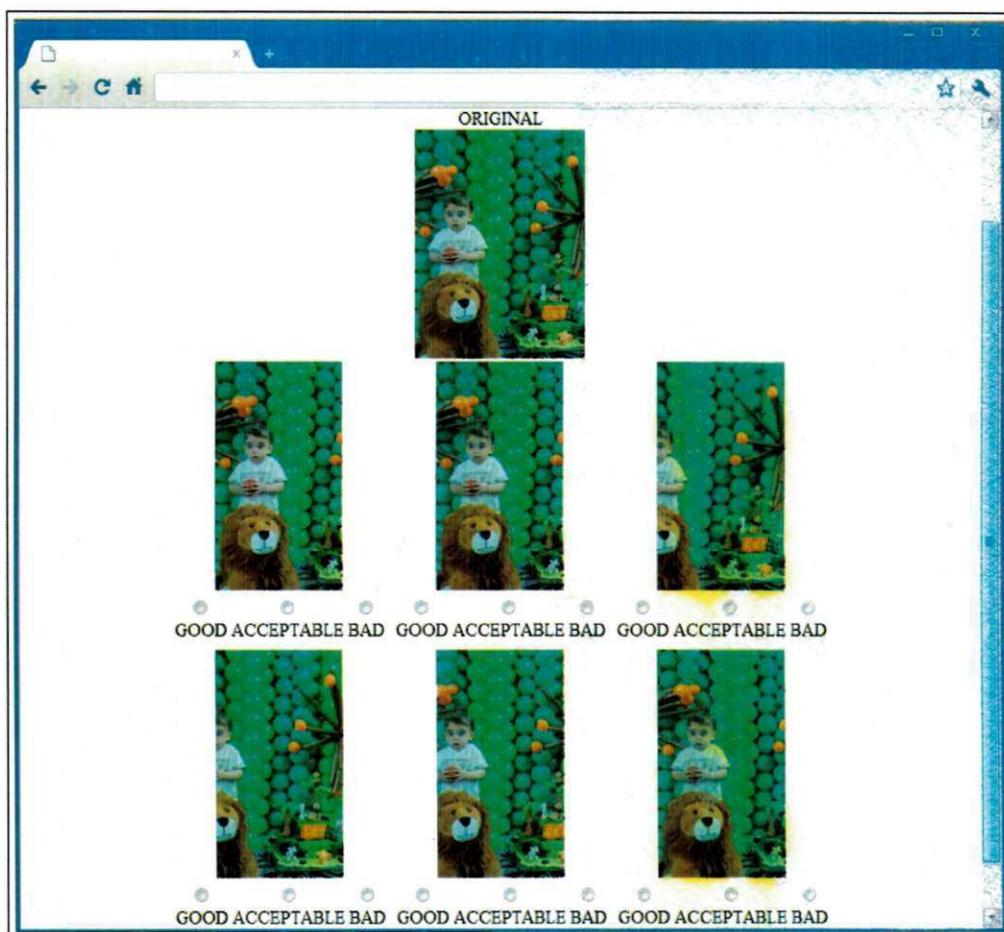
Apesar do principal objetivo do método investigado ser a realização do recorte automático de fotografias, algumas imagens que não continham alvos humanos foram intencionalmente mantidas no conjunto, a fim de se pudesse verificar a robustez do algoritmo no tocante a imagens desta natureza.

Cada uma das 100 fotografias foi apresentada, em uma página *Web*, a um conjunto de 12 participantes, tendo sido a imagem original posicionada no topo da página, enquanto as outras (cada uma referenciando um diferente mapa de saliência, mais o método AZ&C) foram posicionadas em duas linhas, cada uma contendo três imagens. A posição das imagens resultantes na tela foi aleatoriamente modificada, de maneira que o posicionamento do método não pudesse ser memorizado pelo participante. Esta estratégia teve por objetivo evitar que um determinado método que tivesse obtido a preferência do usuário em exemplos anteriores o fizesse repetir a marcação positiva apenas com base no histórico, sem a devida observação.

Próximo a cada imagem, três âncoras semânticas foram apresentadas: “Aceitável”, “Boa” e “Ruim”. Foi solicitado aos participantes que classificassem cada uma das seis imagens a partir de uma das opções apresentadas. Também foi solicitado que a análise fosse realizada a partir da observação de cada fotografia, individualmente. Apesar da segunda solicitação, parecia natural que os observadores efetuassem comparações. Por exemplo, mesmo que um corte pudesse ser considerado “Bom” por um dado usuário, como existiam outras imagens para comparação, o usuário poderia ser levado a classificar a imagem como “Aceitável”. Tal estratégia permitiu a verificação do desem-

penho individual de um dado algoritmo, assim como a identificação de métodos tão bons quanto o primeiro, de acordo com o julgamento subjetivo dos participantes. O formulário *Web* utilizado encontra-se na Figura 7.3.

Figura 7.3: Página *Web* empregada para capturar a opinião dos participantes sobre a qualidade dos recortes realizados.



Uma vez que a votação supramencionada empregou uma Escala de Likert de três pontos (LIKERT, 1932) com âncoras semânticas *Bom*, *Aceitável* e *Ruim*, estas foram mapeadas para os equivalentes numéricos 2,1 e 0 respectivamente. Cada um dos seis métodos teve seus valores numéricos totalizados para todas as imagens e todos os observadores humanos. Uma vez que 12 rotulagens foram consideradas por fotografia, 100 fotografias e o escore para cada uma destas fotografias poderia variar entre 0 e 2, sendo o escore máximo atingível por um dado método igual a $12 \times 100 \times 2 = 2400$, o qual foi utilizado para normalizar as somas absolutas em percentuais.

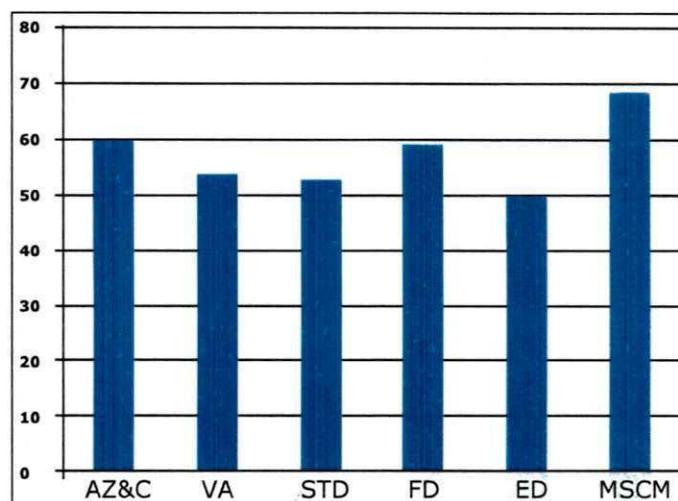
A soma e percentual relativo a cada um dos métodos são apresentados na Tabela 7.4.

Tabela 7.4: Método da Somatória de Valores (LIKERT, 1932) aplicado aos resultados da pesquisa envolvendo 100 imagens e 12 participantes.

AZ&C	VA'	STD'	FD'	ED'	MSCM
1438	1292	1267	1418	1199	1647
59,9%	53,8%	52,8%	59,0%	49,9%	68,6%

Os percentuais da Tabela 7.4 são ilustrados no gráfico em barras apresentado na Figura 7.4, de acordo com o *Método da Somatória de Valores* (LIKERT, 1932).

Figura 7.4: Gráfico em barras ilustrando a preferência dos métodos (em percentuais)



A mesma análise apresentada na Tabela 7.4 foi realizada sem os 6% de alvos não-humanos. Os resultados são apresentados na Tabela 7.5.

Tabela 7.5: Método da Somatória de Valores (LIKERT, 1932) aplicado aos resultados da pesquisa envolvendo 94 imagens e 12 participantes.

AZ&C	VA'	STD'	FD'	ED'	MSCM
1320	1199	1182	1344	1114	1558
58,5%	53,1%	52,4%	59,6%	49,4%	68,1%

Pôde-se perceber que os percentuais pouco variaram, havendo uma variação um pouco maior nos resultados do método detector de faces (0,6%), o que é um indício de

que a presença de imagens contendo alvos não-humanos foi corretamente tratado pelos métodos utilizados.

Quando comparando com as características individuais, o método também foi o escolhido na maioria dos casos. Resultados da comparação direta são mostrados na Tabela 7.6. Em tal tabela, cada linha representa o número de vezes que o método foi considerado superior por um dado usuário com relação ao método indicado na coluna, para uma mesma imagem. Não estão representados nesta tabela os casos em que os métodos foram considerados similares.

Tabela 7.6: Comparação direta entre os métodos. Cada linha mostra o número de vezes que o recorte realizado por um método foi considerado superior ao recorte realizado pelo método indicado na coluna.

	AZ&C	VA'	STD'	FD'	ED'	MSCM
A.Z&C	-	399	417	358	402	237
VA'	290	-	297	337	314	208
STD'	269	298	-	301	322	155
FD'	311	421	386	-	388	135
ED'	216	249	288	268	-	131
MSCM	375	450	408	304	431	-

Em uma comparação direta com o método AZ&C, o MSCM foi preferido pelos participantes em 61% dos votos (375 dentre 612). Quando considerando o método igual ou superior, o MSCM obteve 80% da preferência (963 dentre 1200).

Similarmente, são apresentados na Tabela 7.7 os resultados quando o método foi considerado igual ou superior pelo usuário.

Tabela 7.7: Valor absoluto e percentual de votos nos quais a imagem gerada pelo MSCM foi considerada igual ou superior às imagens geradas pelos métodos empregando características individualmente.

VA'	STD'	FD'	ED'
992	1045	1065	1069
83%	87%	89%	89%

A partir da análise dos resultados, constata-se que o FD' foi considerado o método que menos diferiu do MSCM. Esse resultado era previsto, uma vez que a característica FD apresenta maior peso na combinação das características (vide Tabela 6.1). Entretanto, também pôde ser concluído que apenas o método FD', isoladamente, não produz resultados melhores que o método AZ&C (311x358, vide Tabela 7.6). Tal resultado também era esperado, devido ao fato de muitas imagens não conterem alvos humanos ou estes não terem sido corretamente detectados. Na maioria das imagens que não continham alvos humanos, o FD' foi considerado o pior método. Assim, é possível concluir que a combinação das características tem um papel importante na produção de um recorte melhor na comparação com o método AZ&C. Investigações futuras incluem a repetição deste experimento considerando combinações de características em pares e trios.

O teste ANOVA fator único foi administrado, a fim de verificar se, de acordo com o julgamento humano, existiam diferenças significativas entre os métodos testados. Os dados usados para análise de cada imagem foram escores s'_i , obtidos a partir dos rótulos mapeados, ou seja:

$$s'_i = \sum_{j=1}^{12} v_{i,j}, \quad (7.2)$$

na qual $v_{i,j}$ representa o voto da imagem i pelo participante j . Portanto, cada método comparado produziu 100 escores (um por imagem), os quais foram avaliados a partir do teste ANOVA.

Com 95% de confiança, a análise mostrou que os métodos apresentaram diferenças. Para identificar quais os métodos que apresentaram diferenças estatisticamente significativas, o teste de Tukey-Kramer para comparações pareadas (WALPOLE et al., 2008) foi também administrado.

Para o experimento realizado, concluiu-se que o MSCM não apresentou resultados estatisticamente diferentes do FD' e do AZ&C, mas apresentou resultados estatisticamente diferentes dos demais métodos.

Tal resultado indica que não existe certeza da superioridade do método MSCM sobre AZ&C e o FD', apesar de ser possível afirmar que o MSCM é superior a todos

os demais métodos testados.

Um sumário do teste de Tukey-Kramer de comparações pareadas é apresentado na Tabela 7.8.

Tabela 7.8: Resultados do Teste de Comparação Pareada de Tukey-Kramer entre o MSCM e os métodos que utilizam características individuais.

Comparação pareada	Diferenças significativas
MSCM vs ED'	SIM
MSCM vs STD'	SIM
MSCM vs VA'	SIM
Todas as Outras Combinações	NÃO

Entretanto, não é possível afirmar que o MSCM, o AZ&C e o FD' são equivalentes. Conforme mostrado na Tabela 7.9, o agrupamento realizado pelo teste de Tukey-Kramer denota superposição dos métodos AZ&C e FD com outros métodos que apresentaram escores inferiores.

Tabela 7.9: Agrupamento realizado pela aplicação do Teste de Tukey-Kramer.

Grupo	Métodos
A	MSCM
A & B	AZ&C, FD'
B	VA', STD', ED'

Investigações futuras são necessárias para identificar as razões de tal agrupamento pelo teste de Tukey-Kramer. Uma possibilidade reside na existência de um grande número de imagens similares produzidas mesmo com a aplicação de métodos diferentes. Tal semelhança pode ser consequência de algum viés na base de imagens, ou seja, a existência de casos muito fáceis (para os quais todos os métodos se comportam bem) ou muito difíceis (para os quais existem poucas ou nenhuma solução e todos os métodos se comportaram mau). Outras possibilidades são (i) os métodos são equivalentes e (ii) é difícil para os participantes perceberem diferenças sutis entre os métodos.

7.2 Experimento para a obtenção e validação de um segmentador de imagens orientado às cores

Esta seção foi dividida em duas partes. Na primeira (Seção 7.2.1), discute-se o resultado do processo de obtenção do segmentador por cor, enquanto na segunda (Seção 7.2.2), discute-se a eficácia deste segmentador para a obtenção de cores comparado às cores selecionadas por voluntários.

7.2.1 Resultados preliminares da segmentação de imagens

Um treinamento foi realizado com os algoritmos mencionados na Seção 4.2.1. Foi definido 26 como o número mínimo de espécies (duas a mais que o número de núcleos do *cluster* no qual o experimento foi executado) e que o algoritmo seria interrompido na centésima era. Também foi definida empiricamente, com base em execuções preliminares, a população mínima inicial de 100 indivíduos por espécie (para aumentar a cobertura do espaço de busca nas espécies de maior complexidade), evoluindo cada população por 20 gerações. A taxa de substituição foi de 40%, a taxa de mutação foi de 5% e a probabilidade de cruzamento foi de 70%. Estes valores foram definidos empiricamente de modo que a busca pela solução pudesse ser mais agressiva, dado o baixo número de indivíduos e gerações. O baixo número inicial de indivíduos teve por objetivo evitar perda de tempo com espécies pouco promissoras. À medida que um ciclo de evolução ocorria, espécies mais promissoras tinham o número de indivíduos gradualmente aumentado, conforme explicitado na Equação B.12.

Para o processo de otimização, foram utilizadas 17 imagens. Para cada imagem, foi gerada manualmente uma imagem de referência (*groundtruth*). Quatro voluntários prepararam estas imagens de referência. Cada imagem foi gerada por um único voluntário. Foi solicitado a cada voluntário que indicasse, para cada região da imagem, uma cor que, em sua opinião, representasse aquela região. A definição das fronteiras entre as regiões foi realizada pelo voluntário. Quanto ao nível de segmentação, este foi deixado a caráter do usuário. Entretanto, foi fornecida a recomendação de segmentar

uma região apenas se nela existissem cores que se destacassem das demais. Para evitar muitos estilos de segmentação, foi fornecido aos voluntários um exemplo de segmentação em que as regiões de cor eram segmentadas à medida que na região havia drástica alteração nas cores.

A função objetivo f_{obj} definida para este treinamento consistiu em produzir uma imagem I , sendo cada região segmentada $R|R \subset I$ colorida com a cor c que melhor representasse aquela região, conforme a Equação 7.3, na qual $e(a, b)$ corresponde à distância euclidiana entre as cores a e b .

$$c = \arg \min_{c_j} \sum_{i=1}^{|R|} e(c_i, c_j) \quad c_i, c_j \in R \quad (7.3)$$

Uma vez gerada a imagem, esta foi comparada com a imagem de referência, utilizando-se o espaço de cor Lab, uma vez que as distâncias numéricas são proporcionais à distância de percepção. Foi calculada a diferença *pixel a pixel*, em cálculo análogo ao da Equação 4.2.

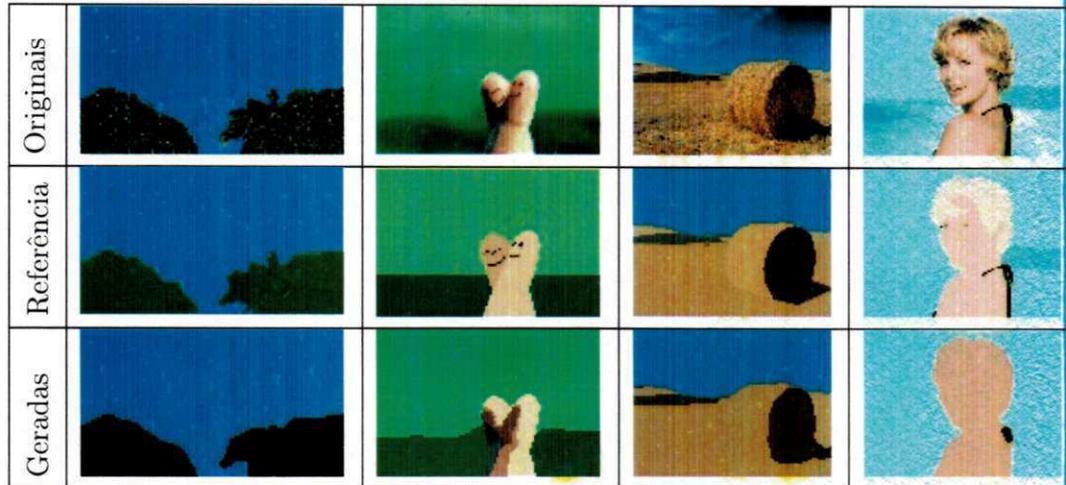
A espécie vencedora deste experimento, após a execução do processo de evolução por 18 eras cada uma dispendo de 20 gerações, constituiu-se da seguinte sequência de métodos (os valores entre parênteses, quando existirem, correspondem aos parâmetros da função): (1) Entrada RGB, (2) Conversão RGB -> HSV, (3) Filtragem homomórfica, (4) Conversão HSV -> RGB, (5) Suavização, (6) MeanShift e (7) Saída RGB.

A sequência encontrada permite inferir que, apesar de ter sido aquela que, dentre as 112 testadas, obteve melhores resultados, ainda há espaço para melhoramentos e otimizações. Inicialmente, após a conversão da imagem para o espaço de cor HSV, há a filtragem homomórfica, com parâmetros de frequência de corte igual a 771, ganho igual a 2 e ordem igual 0. Em seguida, após nova conversão para o espaço de cor RGB, realizou-se uma suavização da imagem, a partir do filtro da mediana de dimensões 3x3. Por fim, foi utilizado o algoritmo MeanShift, com σ_s igual a 12, σ_r em 9.82 e mínimo de 30 regiões.

Algumas das 17 imagens utilizadas durante o treinamento, suas respectivas imagens de referência (*groundtruth*) e imagens correspondentes geradas pelo algoritmo de

segmentação obtido são mostradas na Figura 7.5.

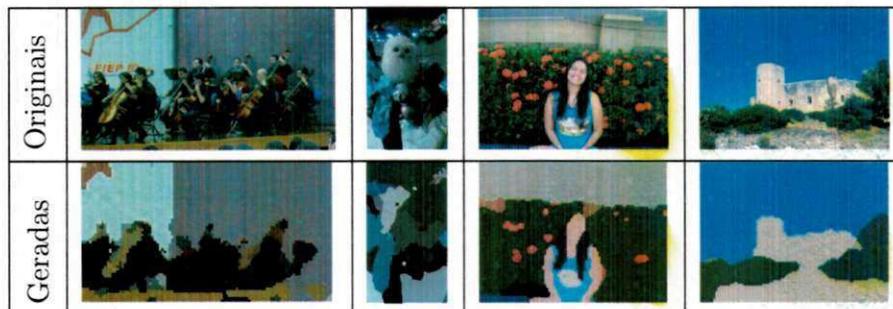
Figura 7.5: Exemplos de imagens da base de treinamento, as respectivas imagens de referência e as imagens geradas pelo sistema resultante do método proposto.



Nas imagens geradas pelo algoritmo de segmentação, cada região foi colorida com a cor pertencente àquela região que apresentou menor diferença com relação às demais cores. As demais imagens encontram-se no Apêndice F.

Na Figura 7.6, são mostradas as imagens de teste e as imagens geradas após a execução da combinação de métodos obtida ao final da otimização. As imagens geradas são em baixa resolução (100x75).

Figura 7.6: Exemplos de imagens da base de testes, geradas a partir da espécie vencedora.



Pode-se perceber, a partir da Figura 7.6 que, de certa forma, o objetivo foi atingido, pois a complexidade das imagens foi reduzida e os principais núcleos de cor foram preservados. Em algumas imagens de nível de complexidade maior (como as duas

primeiras da esquerda para a direita), a redução da complexidade resumiu regiões que necessitariam de um número maior de segmentos.

Uma vez que o objetivo da administração deste método foi a verificação da harmonia das cores de uma imagem, foram extraídas as três principais cores da imagem segmentada. As cores foram obtidas a partir da quantidade de pontos das regiões segmentadas. Portanto, as três regiões com maior número de pontos tiveram suas cores extraídas e exibidas por ordem de número de pontos. As cores extraídas pelo método proposto são posicionadas lado a lado com as cores extraídas pelo método OSCAR (YAO et al., 2011), a partir de uma aplicação demonstrativa de *software*, disponível *online*⁸, sendo exibidas no Apêndice F.

Pretende-se testar, em investigações futuras, outras abordagens de seleção das cores, e.g., utilizar a cor das regiões com maior número de conexões (sendo uma conexão definida quando duas regiões possuem *pixels* 8-conectados), ao invés de considerar a região com maior número de pontos.

Tendo em vista que a revisão da literatura não permitiu encontrar nenhuma forma direta e objetiva de comparar estes resultados, realizou-se uma comparação subjetiva dos resultados com a participação de voluntários. Tal metodologia de avaliação é apresentada na Seção 7.2.2.

7.2.2 Método para a avaliação do segmentador de cores

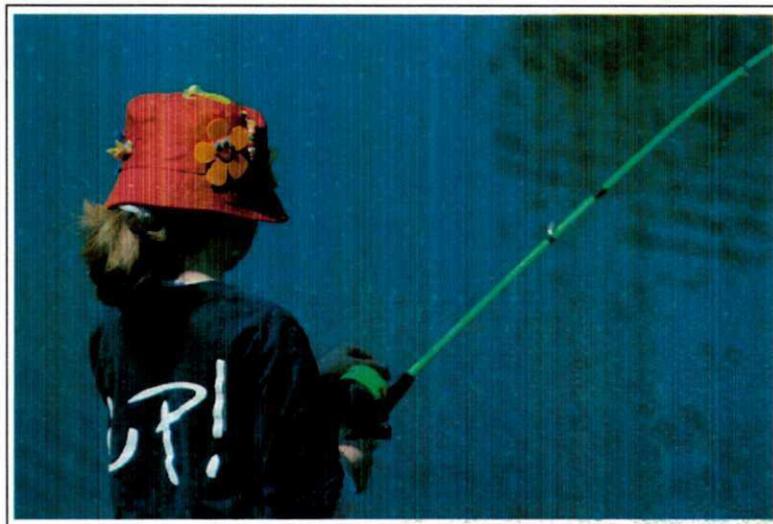
Para avaliar o método de segmentação por cor proposto no Capítulo 4, foi desenvolvido um experimento subjetivo de obtenção das principais cores componentes de uma imagem.

Para que uma análise de cores seja realizada sobre uma imagem, faz-se mister a obtenção de um conjunto de imagens de referência (*groundtruth*) de modo a utilizá-lo para a avaliação de desempenho do método proposto para tal análise. No âmbito desta pesquisa, não foram encontradas bases de imagens públicas indicando quais cores de uma imagem são consideradas mais significativas para a representação de tal imagem para avaliadores humanos.

⁸<http://acquine.alipr.com/oscar/>

Sendo assim, foi desenvolvido um experimento com o objetivo de produzir tal base a qual, posteriormente, pode ser utilizada em outras pesquisas. Este experimento destinou-se à obtenção dos resultados desta pesquisa, uma vez que se desejava verificar o quão as cores de uma imagem são harmônicas. Pode-se adotar uma abordagem global, como aquela proposta por Cohen-Or et al. (2006). Entretanto, a análise local das tonalidades não é suficiente para identificar regiões pequenas de cor mas que possuem grande impacto na percepção visual da imagem, a exemplo do tema da Figura 7.7.

Figura 7.7: Apesar da região vermelha ter área pequena, comparado com o restante da imagem, possui grande impacto na percepção.



A realização deste experimento também foi essencial, dado que a identificação das principais cores da imagem poderia variar para observadores diferentes e, até mesmo, para um mesmo observador. A consequência seria, pois, a discordância que a rotulagem realizada por um único observador poderia apresentar com relação à forma como outros observadores percebem visualmente uma mesma imagem.

Obtenção das rotulagens

Para a obtenção dos dados a serem posteriormente utilizados como *groundtruth*, foi desenvolvido um experimento para a obtenção da opinião de voluntários. O experimento consistiu em apresentar uma imagem ao voluntário e solicitar que este aponte as cores

consideradas por ele como as mais relevantes daquela imagem.

De modo a garantir a abrangência do experimento, foram obtidas 100 imagens de paisagens, objetos e, em maior número, alvos humanos. Detalhes sobre as 100 imagens utilizadas no experimento são mostrados no Apêndice E.

A rotulagem das imagens é um processo lento, uma vez que requer atenção do participante. Para que o número de participantes pudesse atingir um patamar de significância estatística, decidiu-se desenvolver um sistema *Web* utilizando-se a tecnologia GWT, cujos motivos de seleção são explicados na Seção A.3.

A lentidão normalmente associada a processos de avaliação subjetiva explica o número pequeno de imagens utilizadas no experimento. Uma vez que deseja-se o maior número possível de participantes que possa participar do experimento, é necessário que o número de imagens não seja muito grande, o que poderia desestimular o votante ou fazer com que a participação fosse acelerada, comprometendo a participação.

O experimento foi realizado no período de abril a outubro de 2012. A página do experimento está acessível a partir da URL <http://color-analyzer.appspot.com/>. No início do experimento, realizou-se uma coleta de dados sobre o perfil dos participantes. Tal coleta é descrita no Apêndice D.

Validação

Dispondo-se dos resultados da rotulagem, pode-se formular as seguintes hipóteses com relação à percepção de cores dos participantes.

Hipótese 1:

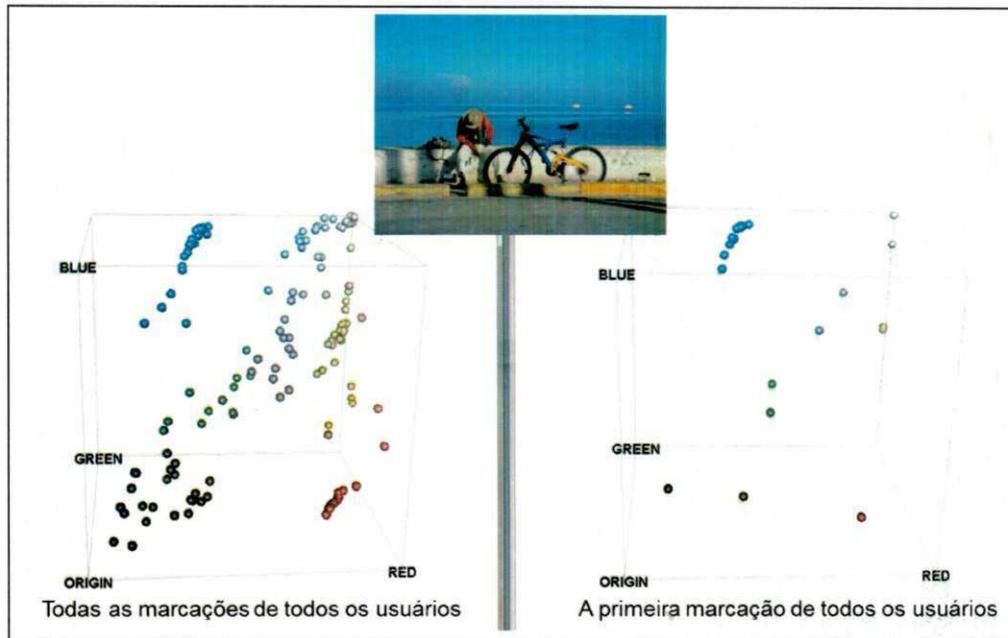
H0 - Os participantes consideram a importância das cores em uma fotografia de forma similar.

H1 - A ordem da importância das cores varia de indivíduo a indivíduo.

Para avaliar esta hipótese, resolveu-se traçar um gráfico no qual as cores são posicionadas no cubo RGB de acordo com a ordem da votação para um participante.

Tomando-se por exemplo a Figura 7.8, referente à rotulagem da fotografia em destaque:

Figura 7.8: Cubo de cores indicando a cor considerada mais importante por todos os participantes do experimento.



Na Figura 7.8, são exibidas, à esquerda, todas as marcações realizadas na fotografia por todos os participantes do experimento. À direita, exibem-se as primeiras marcações (as mais importantes) realizadas por todos os usuários.

Pode-se perceber que os participantes rotularam a cor mais importante de forma bastante divergente. Este resultado pôde ser comprovado tanto visualmente, observando-se o mesmo gráfico de outras imagens, quanto numericamente, a partir do cálculo do erro do centroide, após a execução do algoritmo de *k-means* (BISHOP, 2006) com $k=1$.

Devido a essa divergência quanto à ordenação, decidiu-se dar menos importância à ordem das fotografias e mais importância às cores em si.

Avaliação do método proposto

O método proposto para a obtenção de um número k de cores em uma fotografia digital foi comparado aos seguintes métodos:

- OSCAR (YAO et al., 2011)

- Adobe Kuler (INC., 2013a)
- *k-means* (BISHOP, 2006)

Uma vez definido o k , calcularam-se os k centróides para cada uma das imagens de teste a partir das rotulagens realizadas no experimento supra-mencionado. Os resultados se repetiram para 3 ou 5 centróides uma vez que o método OSCAR (YAO et al., 2011) só permite 3 núcleos e o método Kuler (INC., 2013a) só permite 5 núcleos. O método proposto, por sua vez, possui flexibilidade quanto ao número de núcleos, permitindo seja obtido desde um único núcleo até o número de núcleos equivalente ao número de cores da imagem.

Para cada um dos métodos anteriormente descritos, calculou-se a diferença média entre os centróides calculados por cada método e os centróides calculados pelo método *k-means* (BISHOP, 2006) sobre a base rotulada. Após a realização dos experimentos, verificou-se que o método proposto possui menor diferença para os centróides obtidos pela rotulagem humana do que os métodos testados. Adicionalmente, verificou-se que o resultado é estatisticamente significativo, pela aplicação do teste-T (WALPOLE et al., 2008), com 95% de confiança. O teste-T foi utilizado uma vez que deseja-se comparar diretamente dois métodos.

7.3 Método para a identificação de fotografias similares pelo conteúdo

A identificação de fotografias similares é essencial para uma boa diagramação. Pode-se utilizar o horário das fotografias como principal informação para esta análise. Entretanto, dois problemas impedem a utilização efetiva de tal método: (1) nem todas as fotografias possuem a data e hora em que foram obtidas (ou apresentam erro na marcação destes parâmetros) e (2) não há uma necessária correspondência entre a similaridade e o horário em que a foto foi obtida.

Para resolver este problema, propôs-se um sistema para a avaliação da similaridade de uma fotografia a partir do conteúdo das imagens. Tal abordagem apresenta a vantagem de não permitir a correta identificação da cronologia e de não ser capaz de

identificar todos os alvos existentes na imagem e seus relacionamentos. Dois métodos foram testados para este fim. O primeiro, utiliza histogramas em três dimensões no espaço de cor HSV para comparar as imagens e foi descrito na Seção 5.1.2, enquanto o segundo, utiliza descritores locais para realizar tal comparação (ZHAO; WU; NGO, 2010; ZHAO; NGO, 2013).

Para a comparação entre os métodos, sete bases de fotografias, obtidas em eventos diferentes, tiveram suas fotografias rotuladas duas a duas, em duplicatas ou não. Ao todo, 1.600 fotografias foram utilizadas neste experimento.

Para realizar a comparação, foram utilizados histogramas em três dimensões com 3, 6 e 12 *bins* em cada dimensão. Variou-se também o número de regiões em que a imagem seria dividida, indicada pela quantidade de divisões por orientação (de 1 a 6 divisões). Logo, uma imagem com 3 divisões, possui 9 regiões ($9 = 3 \times 3$). As regiões servem para que a comparação dos histogramas seja feita localmente ao invés de globalmente. Para realizar a comparação entre os histogramas com N bins, foram testados os seguintes métodos: correlação (Equação 7.4), chi-quadrado (Equação 7.5), intersecção (Equação 7.6), Bhattacharyya (Equação 7.7), definidos pelas equações a seguir, e a *Earth's mover distance* (EMD) (RUBNER; TOMASI; GUIBAS, 2000).

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \bar{H}_1)(H_2(I) - \bar{H}_2)}{\sqrt{\sum_I (H_1(I) - \bar{H}_1)^2 \sum_I (H_2(I) - \bar{H}_2)^2}} \quad (7.4)$$

$$d(H_1, H_2) = \sum_I \frac{(H_1(I) - H_2(I))^2}{H_1(I)} \quad (7.5)$$

$$d(H_1, H_2) = \sum_I \min(H_1(I), H_2(I)) \quad (7.6)$$

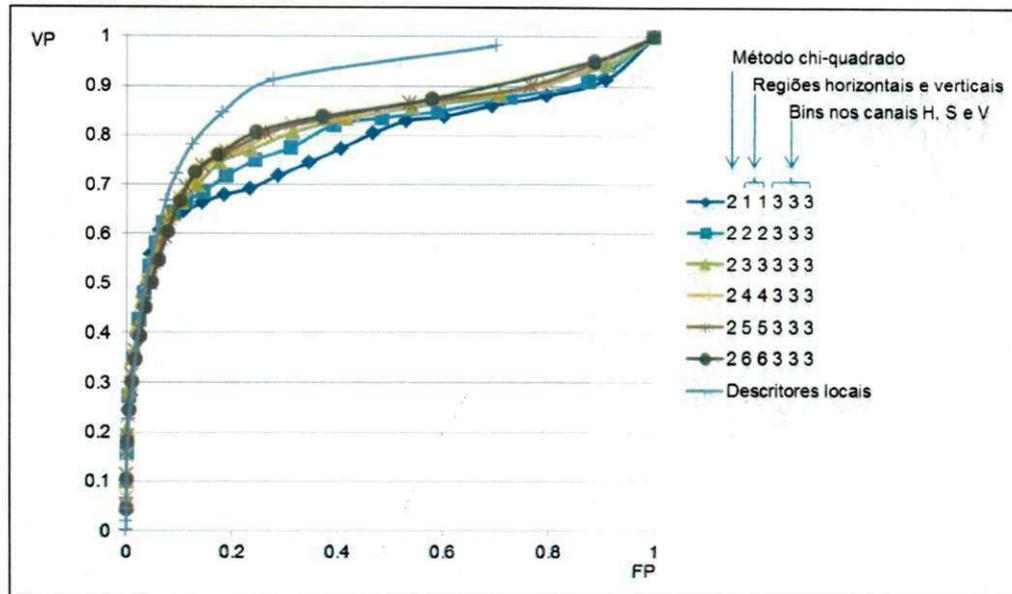
$$d(H_1, H_2) = \sqrt{1 - \frac{1}{\sqrt{H_1 H_2 N^2}} \sum_I \sqrt{H_1(I) H_2(I)}} \quad (7.7)$$

O EMD foi o método que apresentou melhores resultados, seguido pelo chi-quadrado. Entretanto, devido ao tempo necessário para o processamento do EMD, optou-se por utilizar o método chi-quadrado. O tempo necessário para processamento do EMD varia exponencialmente com a quantidade de bins do histograma.

Para comparar os resultados, foi utilizada uma curva ROC (*Receiver Operating Characteristic*). Para visualizar a variação dos resultados na curva, variaram-se os

limiares da ROC para o método chi-quadrado entre 1.000 e 20.000. A curva ROC produzida é exibida na Figura 7.9.

Figura 7.9: Curvas ROC



Na Figura 7.9, cada curva equivale a uma configuração de número de bins e número de regiões da imagem, sendo tal configuração codificada na legenda: 2 1 1 3 3 3, por exemplo, equivale ao método 2 (*chi-quadrado*) com 1 divisão horizontal e 1 vertical e com 3 *bins* nos histogramas do *h*, *s* e *v* respectivamente. Na mesma legenda, Local indica a curva do método que utiliza descritores locais. Pode-se perceber, pelo gráfico exibido, que há muitas semelhanças entre os resultados do primeiro método, mesmo com as variações no número de *bins* e de regiões. De forma geral, percebe-se a melhora dos resultados com o aumento do número de regiões.

Entretanto, ao se comparar os resultados do primeiro e do segundo métodos, o qual utiliza descritores locais (ZHAO; WU; NGO, 2010; ZHAO; NGO, 2013), percebe-se uma vantagem do último. Apesar de tal vantagem, em decorrência do tempo de processamento necessário para analisar características locais, a utilização da comparação via histograma de cor, torna-se mais viável para um problema da natureza do problema tratado nesta pesquisa, o qual requer rápida resposta do sistema ao usuário.

7.4 Avaliação dos métodos para a diagramação de álbuns fotográficos

A etapa final, o módulo diagramador, apresentado no Capítulo 6, produz um álbum fotográfico, tal e qual inicialmente planejado. Entretanto, é essencial verificar se tal álbum produz resultados melhores do que os métodos existentes para tal finalidade. Tal verificação é realizada nesta seção.

A avaliação de um álbum fotográfico é uma tarefa que exige um alto teor de subjetividade. Não existem, de acordo com a pesquisa bibliográfica realizada nesta tese, métricas objetivas que possam ser utilizadas para a avaliação da qualidade de um álbum fotográfico. Naturalmente, poderiam-se definir métricas para a avaliação, e.g., área total ocupada pelas fotografias em cada página do álbum, entretanto, neste caso, seriam necessários estudos para comprovar que tais métricas de fato correspondem às opiniões humanas. Uma vez que a opinião humana é essencial para que a avaliação apresente tal nível de subjetividade, optou-se pela utilização de voluntários humanos para a realização de uma avaliação dos resultados.

Uma vez optado pela utilização de humanos, é essencial que alguns fatores sejam levados em consideração, de modo a reduzir o efeito causado pelas diferentes interpretações que cada participante pode ter sobre o problema. Desta forma, definiram-se as seguintes restrições para a avaliação da diagramação dos álbuns fotográficos:

- **Formato da avaliação.** Uma vez que a avaliação de um leiaute é uma tarefa, por si só, repleta de subjetividade, a avaliação foi realizada por intermédio da escolha forçada, ou seja, serão apresentadas duas opções dentre as quais o usuário terá que optar pela que mais lhe agrade;
- **Leiaute.** O objetivo desta pesquisa não é identificar o tipo de leiaute que mais agrada, dentre todos os tipos possíveis de leiaute. Assim, ambos os leiautes seguiram a mesma lógica de distribuição das fotografias;
- **Fotografias.** Dependendo das fotografias a serem exibidas, a opinião do usuário

quanto ao álbum gerado pode variar. Assim, ambos os álbuns foram compostos pelo mesmo conjunto de fotografias;

- **Intervenção humana.** Uma vez que a proposta trata de diagramação automática, não poderá haver nenhum tipo de intervenção humana na geração dos álbuns, exceto as fotografias de entrada e o número de páginas desejado. Assim, o álbum final gerado deve ser aquele produzido automaticamente pela ferramenta avaliada.
- **Cronologia.** Uma vez que o relacionamento entre as fotografias é um fator relevante para a decisão de quais fotografias devem aparecer em uma mesma página, a cronologia do evento deve estar disponível para os participantes do experimento.
- **Elementos visuais.** Sempre que possível, as páginas geradas devem ter elementos visuais equivalentes, e.g., a cor do plano de fundo deve ser a mesma para evitar julgamentos distorcidos.
- **Número de páginas dos álbuns.** O número de páginas nos álbuns comparados deve ser o mesmo. Assim, evita-se avaliar esforços diferentes de diagramação.

Com base nas restrições anteriormente apresentadas, é necessário a definição do modelo do experimento. Como já discutido, foram dois os álbuns a serem comparados: o método proposto e um método com estilo de leiaute similar. Uma vez que o método proposto é uma extensão do BRIC (ATKINS, 2008), foi utilizado para comparação um outro método que também utilize o algoritmo BRIC. Para tanto, foram identificadas duas opções possíveis: o *autophotobook*⁹ e o *snapfish*¹⁰, ambos serviços *online* para a diagramação de álbuns fotográficos desenvolvido pelo HP Labs. Uma vez que o método proposto tem por objetivo ter como entrada apenas as fotografias a serem diagramadas e o número de páginas que o álbum deve ter, a abordagem escolhida foi o *autophotobook*,

⁹www.autophotobook.com

¹⁰www.snapfish.com

uma vez que o *snafish* não permite a definição do número exato de páginas no qual as fotografias devem ser diagramadas.

Com relação às fotografias, será utilizado o acervo de um voluntário, o qual doou 144 fotografias obtidas em uma viagem do qual o voluntário participou à cidade de Buenos Aires, na Argentina. Tal acervo foi utilizado pois (1) todas as fotografias foram obtidas por fotógrafos amadores, as quais são o foco desta pesquisa, (2) Não houve filtragem por parte do voluntário, assim fotografias consideradas boas e ruins permaneceram no conjunto, (3) as fotografias foram obtidas num período de uma semana, o qual permite explorar uma grande diversidade de intervalos de tempo interfotografias, (4) podia-se dispor de um roteiro temporal do evento e (5) o número razoável de fotografias, fato que permite avaliar um comportamento de seleção sem a necessidade de requisitos temporais extensos para a participação do experimento, além de facilitar a memória dos participantes quanto aos eventos existentes no álbum escolhido.

Como já definido, não é permitida a intervenção dos participantes na geração do álbum. Uma vez que não se dispõe do sistema gerador do álbum fotográfico, foi necessária a utilização da página *Web* existente, o *autophotobook*. Entretanto, o site em questão não dispõe de interfaces de utilização que pudessem ser utilizadas para automatização do processo. Sendo assim, foi criado um roteiro de passos, os quais deviam ser rigorosamente seguidos por cada participante, evitando assim interações extra sobre os álbuns.

Para que fosse garantida a cronologia do álbum, verificou-se que (1) todas as fotografias utilizadas foram obtidas com a mesma câmera, (2) todas as fotografias possuem os metadados EXIF intactos e (3) não houve alteração da data da câmera no decorrer da viagem.

Com relação aos elementos visuais, o álbum gerado pelo *autophotobook*, por padrão, produz uma capa com a data de geração do álbum e tem cor do plano de fundo preta. Para manter a consistência, ambos os elementos foram reproduzidos no método proposto nesta pesquisa.

Por fim, de modo a garantir justiça nos esforços de geração dos álbuns, foi definido

um número comum de páginas por álbum. Em testes piloto, foi verificado que a quantidade de fotografias selecionadas pelos usuários poderia variar, chegando a algumas dezenas. Assim, definiu-se que a quantidade adequada de páginas deveria garantir que não existissem, constantemente, poucas ou muitas fotografias por página (i.e., mais do que 6 fotografias por página, valor definido ao longo do teste piloto pela identificação, por parte dos participantes, das páginas com pior aspecto visual). Desta forma, concluiu-se que o número de páginas para este conjunto de fotografias deveria ser 16. Desta forma, os álbuns gerados possuem exatamente 16 páginas, no mínimo 1 fotografia por página e no máximo 6 fotografias por página. Assim, o número de fotografias selecionadas para compor o álbum deve estar entre 16 e 96 fotografias. Caso o número de fotografias escolhidas pelo participante não estivesse neste intervalo, uma mensagem era exibida de maneira que o usuário pudesse adequar o número de fotografias aos limite permitidos.

O experimento para comparação dos métodos geradores de álbuns fotográficos foi dividido em três partes. A primeira parte, consistiu em apresentar ao participante o roteiro da viagem, o qual associava as 144 fotografias a um determinado sub-evento por intermédio da simulação de um diário e, em seguida, solicitar que o mesmo escolhesse, dentre o conjunto apresentado, quais as fotografias que ilustram de forma completa, porém sucinta, o evento. A segunda parte, consistiu em acessar a página *Web autophotobook* e, utilizando as fotografias previamente selecionadas pelo participante, seguir algumas instruções de utilização da página que eram apresentadas. Por fim, a terceira parte, consistiu em apresentar dois álbuns, aquele gerado pela página *Web* e aquele gerado pelo método proposto, e solicitar ao participante a responder perguntas comparativas entre os álbuns. As três partes são detalhadas nas Seções 7.4.1, 7.4.2 e 7.4.3 a seguir.

7.4.1 Seleção das fotografias

A etapa de seleção de fotografias iniciou com a apresentação do “diário da viajante”. Tal diário foi criado com o intuito de explicar o conteúdo das fotografias e apresentar

a sequência temporal das fotografias. Esta etapa é necessária uma vez que os participantes podem não conhecer os lugares onde as fotografias foram obtidas nem os personagens principais. Assim, garante-se que haja um critério menos aleatório na seleção das fotografias.

O diário da viajante é apresentado no Apêndice G. Ao todo, 67 participantes realizaram o experimento. Cada participante selecionou as fotografias que considerava mais relevantes para o relato do evento, o que totalizou 67 seleções dentre as 144 fotografias, contendo de 16 a 96 fotografias cada uma.

Para realizar a avaliação, utilizou-se a seguinte abordagem: foi desenvolvido um seletor de fotografias, a partir da combinação das métricas desenvolvidas. A seleção das fotografias tem forte relação com a ligação temporal e a similaridade existente entre as fotografias. Assim, foi utilizado o método para cálculo de similaridade descrito no Capítulo 5. Para cada grupo de fotografias similares, utilizou-se o método seletor descrito no Capítulo 4 para identificar se alguma das fotografias seria selecionada.

De forma a permitir uma comparação entre os métodos supracitados, foi utilizada a própria seleção realizada pelos participantes do experimento. Assim, pôde-se avaliar quão bem um humano realiza esta seleção e, portanto, comparar o método desenvolvido ao desempenho humano.

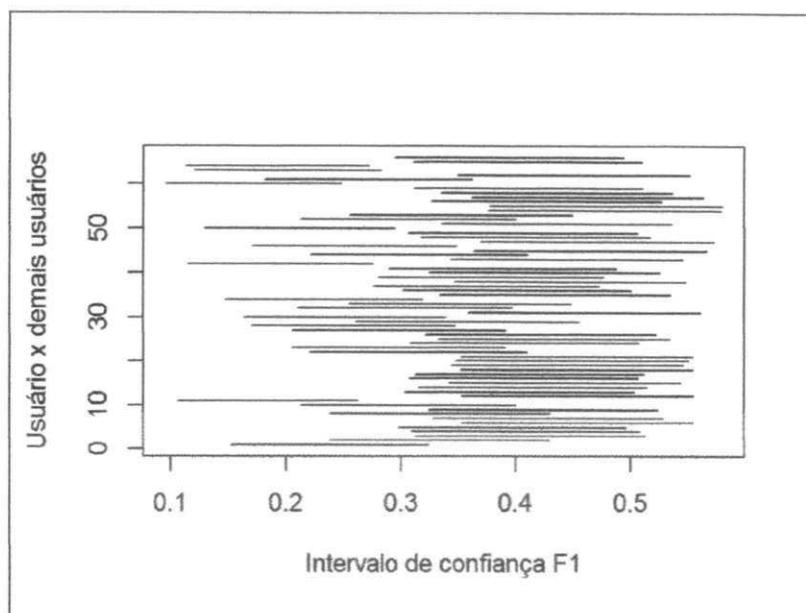
Tal tipo de avaliação é importante para constatar a complexidade da tarefa. Se humanos não forem capazes de realizar uma tarefa de cunho subjetivo de forma consistente, não há como se esperar desempenho muito superior de algoritmos que objetivam reproduzir o comportamento humano.

Para tanto, cada uma das seleções existentes foi comparada a todas as outras seleções. A avaliação, portanto, levou em consideração as fotografias que não foram escolhidas, mas que deveriam ter sido, como também as que foram escolhidas mas que não deveriam ter sido. Por conseguinte, as medidas mais indicadas para esta avaliação são a precisão e a revocação.

Sendo assim, como resultado final, foi calculada a medida-F para cada seleção, uma vez que esta resume em um só valor a precisão e revocação. Assim, compara-se um

usuário com os demais 66. Para cada seleção, portanto, há 66 valores de medida-F. Sumaria-se esses valores a partir do intervalo de confiança a 95%. Os intervalos de confiança da medida-F calculados para cada um dos 67 usuários são ilustrados na Figura 7.10.

Figura 7.10: Intervalos de confiança da medida-F dos resultados de seleção de um usuário para os demais.



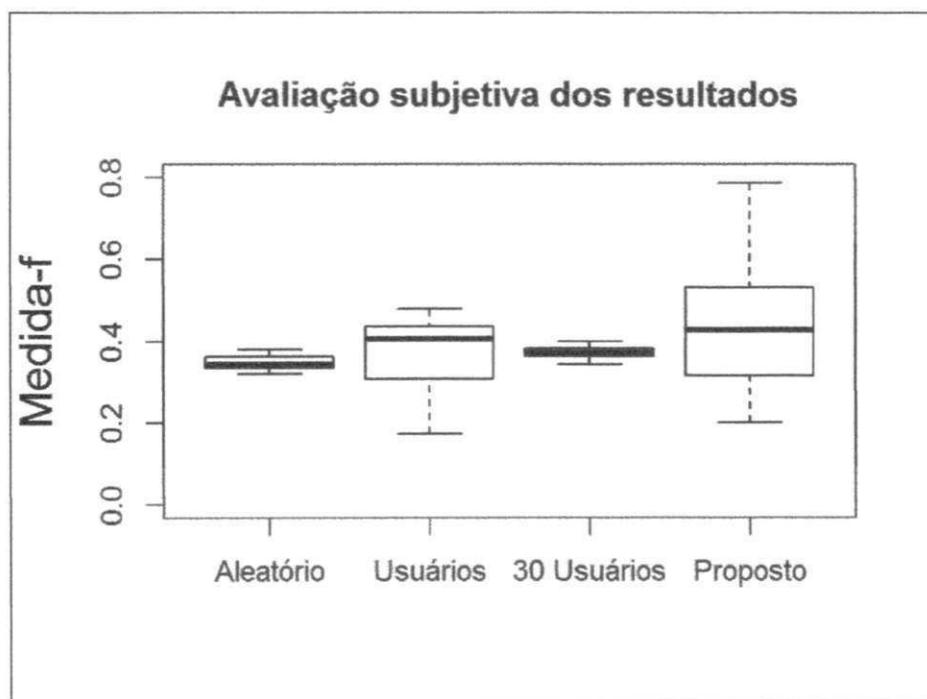
A partir da análise realizada, pode-se perceber o nível de divergência existente entre os participantes dos experimentos. Uma possível explicação é que não há um consenso na seleção de fotografias. Dentre as possíveis hipóteses para este resultado:

- as fotografias possuem um nível similar de qualidade - o que pode ter gerado inconsistência na seleção;
- nem todos os participantes levaram em consideração a qualidade das fotografias - logo, critérios diferentes resultariam em seleções inconsistentes; e
- a seleção das fotografias apresenta inconsistência inerente ao problema.

Uma conclusão que pode ser derivada destas observações é que o método proposto dificilmente obterá resultados significativamente melhores, uma vez que até as observações humanas não obtiveram resultados razoáveis.

Para realizar a comparação, foram testados quatro métodos: (1) seleção aleatória; (2) seleção realizada por todos os participantes; (3) seleção realizada por um grupo aleatório de 30 participantes; (4) o método proposto. A partir dos valores da medida-f de cada um dos quatro métodos, traçou-se um diagrama de caixa (*boxplot*), mostrado na Figura 7.11.

Figura 7.11: Diagrama de caixa com os valores da medida-f dos quatro métodos propostos.



A partir da análise do diagrama de caixa, não é possível identificar qual método é superior. Portanto, foi realizado um teste de hipóteses. Utilizando-se o teste-T, pode-se concluir, com 95% de confiança, que o método proposto (4) é superior ao método de seleção aleatória e equivalente à seleção humana. Com 90% de confiança o método proposto é também superior ao método de seleção por humanos dada uma amostra aleatória de 30 participantes.

A partir dos resultados anteriormente apresentados, decidiu-se explorar variações na abordagem. Para isso, foram removidos do conjunto os usuários que possuíam valor médio da medida-F destoando dos demais. Assim, foram testados os valores para média superior a 0,10, 0,20 e 0,30, mostrados nas Figuras 7.12, 7.13 e 7.14.

Figura 7.12: Diagrama de caixa com os valores da medida-f dos quatro métodos propostos, considerando-se apenas os usuários com a média da medida F superior a 0,10.

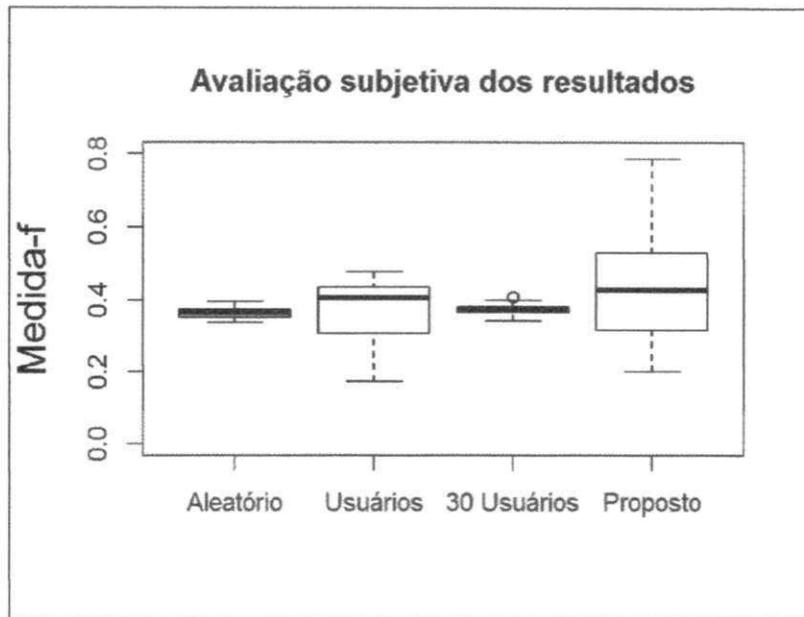
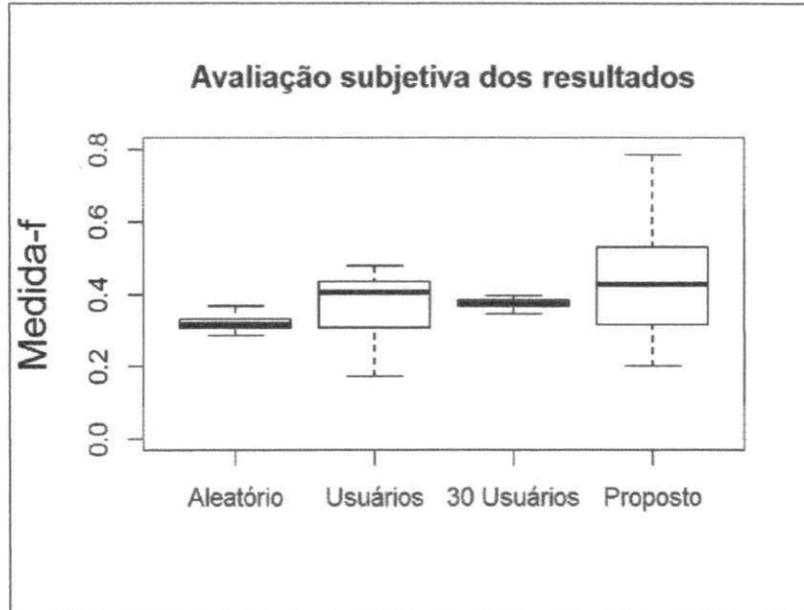


Figura 7.13: Diagrama de caixa com os valores da medida-f dos quatro métodos propostos, considerando-se apenas os usuários com a média da medida F superior a 0,20.



Os diagramas de caixa continuam sem ser suficientes para, isoladamente, indicar o melhor método. Utilizou-se o teste-T (WALPOLE et al., 2008), com 95% de confiança, e concluiu-se que o método proposto é superior aos demais métodos testados.

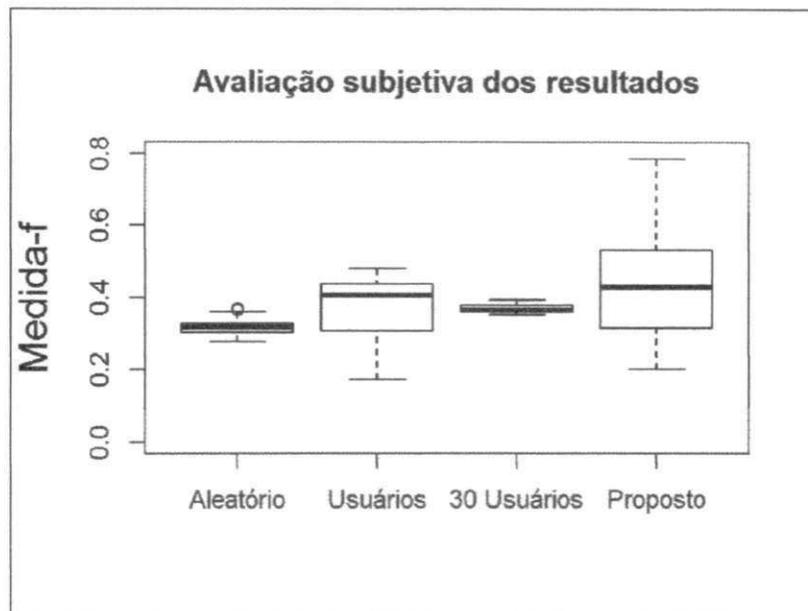


Figura 7.14: Diagrama de caixa com os valores da medida-f dos quatro métodos propostos, considerando-se apenas os usuários com a média da medida F superior a 0,30.

Pode-se concluir que o método desenvolvido para a seleção de fotografias é equivalente à seleção realizada por humanos, apesar da inerente dificuldade de obtenção de melhores resultados para este problema.

7.4.2 Experimento com o *autophotobook*

Para a comparação dos resultados das duas abordagens, foi realizado um experimento em que cada participante seguia uma sequência de instruções para a utilização do site *autophotobook*. Tais instruções orientavam o participante a realizar o *upload* das fotografias selecionadas e gerar o álbum com as características previamente definidas, e.g., o número de páginas em 16.

O experimento foi realizado em dois computadores *HP TouchSmart 410 BR*, especialmente preparados para este experimento. Desta forma, dois participantes poderiam realizar o experimento simultaneamente. Uma vez seguidas todas as instruções, o álbum gerado era obtido e gravado no diretório adequado para a apresentação ao usuário, de modo que a interface produzida para a comparação entre álbuns fosse apresentada corretamente. De modo a não interromper o fluxo do experimento, enquanto as instru-

ções eram seguidas pelos usuários, o sistema proposto era executado em paralelo.

Apesar do experimento ter sido realizado normalmente com a participação de oito voluntários, foi necessário interromper o experimento em decorrência da remoção da página do *autophotobook* do ar no segundo dia de experimento. Uma vez que a quantidade de participantes tornou-se insuficiente para a obtenção de conclusões robustas sobre o experimento e dado que o *autophotobook* não foi restaurado, fez-se necessário uma mudança na abordagem.

Dado que existiam oito álbuns diferentes, decidiu-se utilizar estes álbuns para realizar a comparação. Foram utilizados os oito álbuns para evitar que houvesse algum viés em decorrência da escolha das fotografias por parte de um dos oito participantes que realizaram o experimento. Desta forma, um segundo experimento foi realizado, no qual a etapa de produção do álbum pelo *autophotobook* foi eliminada. Assim, logo após a seleção das fotografias, um dos oito álbuns era escolhido aleatoriamente e exibido ao participante. Dado que as fotos escolhidas pelos usuários não seriam mais exibidas no álbum gerado, foi adicionada a observação de que os álbuns seriam produzidos com as fotografias escolhidas pela proprietária das fotografias, de acordo com o interesse individual da mesma.

Este segundo experimento foi realizado por 67 participantes diferentes dos 8 que participaram da primeira versão do experimento. Uma vez que não houve mais a produção dos álbuns a partir das fotografias selecionadas, decidiu-se por gerar uma versão *online* do experimento, a qual foi hospedada no *Google App Engine* (INC., 2013b). Os 8 álbuns produzidos estão disponíveis a partir do endereço <https://sites.google.com/a/copin.ufcg.edu.br/albumproduzidos/>.

7.4.3 Comparação entre os álbuns utilizados

Uma vez dispondo-se dos álbuns a serem comparados, o álbum gerado pelo método proposto e o álbum gerado pelo *autophotobook*, apresentou-se aos participantes um questionário comparativo entre tais álbuns. Ao todo, dez perguntas foram realizadas, as quais são listadas a seguir:

1. Qual sua impressão sobre o álbum da esquerda?

2. Qual sua impressão sobre o álbum da direita?
3. De maneira geral, de qual dos dois álbuns você gostou mais?
4. Qual dos álbuns organizou melhor as fotografias?
5. Em sua opinião, quais as melhores páginas do álbum da esquerda? Por que?
6. Em sua opinião, quais as piores páginas do álbum da esquerda? Por que?
7. Em sua opinião, quais as melhores páginas do álbum da direita? Por que?
8. Em sua opinião, quais as piores páginas do álbum da direita? Por que?
9. Levando-se em consideração ambos os álbuns, qual o principal problema que você identificou? Por que?
10. Você utilizaria um sistema similar para produzir álbuns com suas fotografias?
Por favor, justifique.

Nesta tese, a análise foi focada nas perguntas 3 e 4, uma vez que estas possuem a análise mais objetiva dos resultados. Assim, com relação à pergunta 3, dentre os 67 participantes, 44 (ou 65%) preferiram o álbum gerado pelo método proposto. Já com relação à pergunta 4, 50 dentre os 67 participantes (ou 75%) preferiram a organização realizada pelo método proposto.

Ao analisar as justificativas dadas pelos participantes com relação às respostas dadas para as perguntas 3 e 4, pode-se perceber que alguns usuários não seguiram as instruções dadas, realizando comentários que iam de encontro aos critérios definidos para o questionário, e.g., uma das respostas indicou que o usuário preferiu um dos álbuns pois achou que o conteúdo das fotografias era mais atraente naquele álbum, quando na verdade ambos os álbuns eram compostos pelas mesmas fotografias, logo, o usuário certamente era inconsistente. Em removendo-se os participantes que realizaram comentários similares, restaram 47 participantes, dentre os quais 36 (ou 77%) optaram pelo álbum produzido pelo método proposto (pergunta 3) e 38 (81%) preferiram a organização realizada pelo método proposto (pergunta 4) ambos em relação ao método concorrente.

Os principais motivos que levaram os participantes a preferir o álbum gerado pelo *autophotobook*, dizem respeito à seleção das fotografias por página. Para tais participantes, a organização das páginas do álbum não deveria ter sido realizada a partir da ordem cronológica, pois (1) fica cansativo visualizar fotos similares e (2) preferem o relacionamento visual do que o temporal.

A grande maioria das justificativas dadas pelos participantes pela preferência ao método proposto está relacionada à organização cronológica das fotografias do álbum.

Capítulo 8

Considerações finais

Ao longo desta pesquisa, foram apresentados diversos métodos destinados à análise e ao processamento de fotografias digitais, tendo em vista a produção automática de álbuns fotográficos.

Foi proposto e desenvolvido um sistema constituído por três etapas: seleção, agrupamento e diagramação. No Capítulo 4, apresentaram-se métodos complementares aos propostos por Cavalcanti (2007), com o intuito de melhorar tal etapa. Após terem sido realizados os experimentos, pôde-se comprovar que o método proposto para a seleção de fotografias tem resultados similares àqueles produzidos por humanos. No Capítulo 5, destinado ao método de agrupamento, apresentaram-se métodos destinados ao agrupamento de fotografias, tanto a partir de suas meta-informações (e.g., utilizando-se a informação de horário) quanto pelo conteúdo da fotografia (i.e., utilizando-se da distribuição das cores existentes). Por sua vez, na etapa de diagramação, apresentada no Capítulo 6, utilizou-se um arcabouço existente, o BRIC, de modo a ampliar a utilização do espaço disponível a partir do recorte e da redistribuição das fotografias.

De forma geral, pode-se afirmar que esta é uma área de pesquisa que apresenta diversas dificuldades, uma vez que é necessário o desenvolvimento de algoritmos cujos resultados sejam similares àqueles obtidos a partir do processamento humano. Contudo, é necessário se destacar que, em muitas oportunidades, nem mesmo os resultados dos humanos apresentam consistência para outros humanos, em decorrência do alto grau de subjetividade associado, como se pôde observar a partir dos resultados dos experimentos, apresentados no Capítulo 7.

Assim, esta pesquisa não deve ser tida como uma investigação definitiva na área, mas sim o estudo, a proposição e a integração de um conjunto de abordagens promissoras com vistas a avançar na direção de melhor compreensão de alguns aspectos da opinião humana (e.g., no que diz respeito à organização de coleções de fotografias).

Apesar de todas as dificuldades ora elencadas, podem-se destacar algumas contribuições advindas da realização desta pesquisa, as quais são listadas na Seção 8.1. Parte destas contribuições já foram publicadas em veículos, tais como conferências e periódicos, os quais são listados na Seção 8.2. Na Seção 8.3, apresentam-se as limitações da abordagem proposta. Por fim, na Seção 8.4, apresentam-se sugestões de investigações que podem dar continuidade à pesquisa desenvolvida nesta tese.

8.1 Contribuições

Pode-se destacar, como principal contribuição desta pesquisa, o sistema desenvolvido para a diagramação automática de álbuns fotográficos, o qual pode ser utilizado com o intuito de sumariar um conjunto de fotografias, sem a necessidade de intervenção humana.

Adicionalmente a esta contribuição, podem ser elencadas outras contribuições desta pesquisa:

- Melhorias nos métodos destinados à diagramação de fotografias;
- Revisão do estado-da-arte de métodos destinados à análise de fotografias, agrupamento de fotografias e diagramação de álbuns fotográficos;
- Proposição de regra de composição fotográfica destinada à análise do espaço negativo;
- Método para a segmentação automática de fotografias em regiões de cores representativas de uma imagem;
- Método para a aquisição de operadores complexos de processamento de imagens (e.g., a segmentação automática por cores, comentada na Seção B.1.5), a partir de uma estratégia de computação evolucionária;

- Método para o agrupamento de cores fundamentado no espaço de cor HSV;
- Método para o recorte automático de fotografias fundamentado no conteúdo, visando à adaptação de fotografias a diferentes proporções-alvo (MSCM);
- Melhorias nos métodos existentes, visando ao agrupamento de fotografias a partir do horário da captura;

8.2 Publicações resultantes desta pesquisa

Parte dos resultados da presente pesquisa encontra-se publicada em alguns veículos nacionais e internacionais da área. A primeira publicação, relacionada à investigação da composição automática de fotografias, mas tendo-se como restrição a existência de um único alvo humano (CAVALCANTI et al., 2010), foi publicada na conferência WSCG (*International Conference on Computer Graphics, Visualization and Computer Vision*).

Em seguida, a investigação referente ao recorte com respeito ao conteúdo foi publicado na conferência SITIS (*Signal-Image Technology & Internet-Based Systems*) (CAVALCANTI; GOMES; QUEIROZ, 2010a).

O terceiro artigo publicado consiste em uma extensão desta última investigação, desta vez publicada no JMPT (*Journal of Multimedia Processing and Technologies* (CAVALCANTI; GOMES; QUEIROZ, 2010b)).

Por fim, uma versão estendida da pesquisa bibliográfica realizada, referente aos métodos destinados à análise e processamento de fotografias digitais, foi publicada no Springer-JBCS (*Journal of Brazilian Computer Society*) (CAVALCANTI; QUEIROZ; GOMES, 2013).

8.3 Limitações da abordagem proposta

Pode-se elencar as seguintes limitações da abordagem proposta, as quais podem ser utilizadas como motivações para investigações futuras:

- O método utilizado para a identificação de fotografias similares por conteúdo

pode apresentar falhas no caso de duas imagens serem diferentes (ainda que em poucos detalhes) mas possuírem distribuições de cor similares em todas as regiões analisadas;

- A abordagem foi desenvolvida e testada a partir de um número limitado de conjuntos de fotografias. Diferentes conjuntos podem, eventualmente, apresentar resultados divergentes;
- Para a correta utilização do módulo analisador de fotografias, é necessário que os elementos analisados da imagem (e.g., faces) sejam corretamente identificados;
- O módulo de agrupamento apresentado limita-se a imagens cujas diferenças cromática e cronológica sejam suficientes para a correta identificação dos grupos;
- Os resultados de diagramação limitaram-se à utilização do algoritmo BRIC, apesar de outros modelos de leiaute podem ser utilizados;
- As análises foram realizadas a partir de um conjunto rotulado por participantes cujos perfis são apresentados no Apêndice D;
- As fotografias utilizadas precisam pertencer a um único evento; e
- É preciso, para o correto agrupamento, que a informação cronológica de obtenção de todas as fotografias utilizadas estejam disponíveis e sincronizadas.

8.4 Investigações futuras

Conforme mencionado anteriormente, apesar dos avanços consequentes desta pesquisa, ainda há muito a ser investigado, a fim de que os problemas relacionados à área pesquisada sejam considerados superados. Sendo assim, há várias sugestões de investigações que podem ser derivadas desta pesquisa.

Uma possível primeira derivação deve tratar da participação do usuário no processo de diagramação de álbuns fotográficos. A produção de álbuns fotográficos pode ser realizada de forma automática, tal e qual apresentado nesta pesquisa, tendo sido estimados e definidos valores padrão para diversos parâmetros (e.g., número máximo

de fotografias por página e diferença máxima das áreas das fotografias permitidas em uma única página). Entretanto, de modo a aumentar a possibilidade de adequação a interesses ou preferências individuais de usuários, pode-se permitir maior intervenção, e.g., um usuário pode preferir ter todas as fotografias em uma quantidade mínima de páginas, enquanto outro pode preferir suas fotografias ocupando o máximo de área possível. Uma vez concluída a produção do álbum, este é exibido ao usuário, o qual pode realizar alterações, caso assim o deseje. Neste caso, tem-se um diagramador semi-automático de fotografias. Pretende-se, portanto, desenvolver em pesquisas futuras um sistema de apoio ao usuário, de modo a facilitar a diagramação de álbuns fotográficos.

Um segundo tema de investigação, que pode ser derivado desta pesquisa, é relacionado ao agrupamento de fotografias. Em virtude do grande número de diferentes possibilidades quanto aos conjuntos de fotografias, não foi possível estudar com profundidade a avaliação e a melhoria dos métodos destinados ao agrupamento. Sendo assim, uma investigação centrada apenas na referida etapa pode trazer melhorias ao processo de produção automática do álbum.

Como terceira sugestão de investigação, propõem-se estudos comparativos entre o método proposto para a obtenção do segmentador por cores e outros métodos canônicos (e.g., programação genética).

Como quarta sugestão de tema de investigação, propõe-se realizar um estudo estendido sobre o impacto da combinação de características no desempenho do algoritmo de recorte automático apresentado na Seção 6.2.2.

Ainda quanto ao algoritmo de recorte automático, pode-se realizar a expansão do algoritmo de recorte, visando a permitir maiores recortes na imagem, uma vez que a abordagem atual recorta o mínimo de área até que a nova proporção seja alcançada. Também é proposta a combinação de mapas de saliência em pares e trios para verificar se existe uma combinação mais simples e equivalente àquela atualmente adotada.

Adicionalmente, propõe-se verificar a influência do algoritmo de recorte no julgamento do votante, assim como um refinamento dinâmico dos pesos a adotar, de modo a atualizá-los à medida que os usuários identificam as imagens cujos recortes não foram satisfatórios (*Relevance Feedback*).

Propõe-se também agregar ao módulo analisador de fotografias outras medidas

de análise, tais como métricas de qualidade (e.g., análise do desfocamento de uma imagem).(COUZINIE-DEVY et al., 2013).

Finalmente, propõe-se como tema final de investigação, o estudo do impacto do layout utilizado na diagramação. Nesta pesquisa, utilizou-se o método BRIC. Entretanto, existem outros algoritmos com a mesma finalidade. Melhor compreensão das preferências dos usuários com relação ao método de diagramação pode garantir a diagramação personalizável aos usuários do sistema.

Referências Bibliográficas

- ACHANTA, R.; SÜSSTRUNK, S. Saliency Detection for Content-aware Image Resizing. In: *Proc. of IEEE ICIP 2009*. Cairo, Egito: [s.n.], 2009. p. 1005–1008.
- ALBATAL, R.; MULHEM, P.; CHIARAMELLA, Y.; CHIN, T.-J. Comparing image segmentation algorithms for content based image retrieval systems. In: *SinFra 2009*. Singapore: [s.n.], 2009. p. 1–10.
- ANGUELOV, D.; LEE, K. chih; GOKTURK, S. B.; SUMENGEN, B. Contextual identity recognition in personal photo albums. In: *Proc. of IEEE CVPR 2007*. Minneapolis, EUA: [s.n.], 2007. p. 1–7.
- ARNHEIM, R. *Arte & Percepção Visual: Uma psicologia da visão criativa*. [S.l.]: Pioneira Thomson Learning, 1980.
- ARRUDA, F. A. P. V. *Geração de representações não foto-realísticas a partir de fotografias digitais: um estudo de caso com faces humanas*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2009.
- ATKINS, B. C. Adaptive photo collection page layout. In: *Proc. of IEEE ICIP*. Singapore, Singapura: [s.n.], 2004. p. 2897 – 2900.
- ATKINS, B. C. Blocked recursive image composition. In: *Proceeding of the ACM MM*. New York, EUA: [s.n.], 2008. p. 821–824.
- ATKINS, C. B.; LYONS, N. P.; ZHANG, X.; TRETTER, D. R. Templated recursive image composition. In: *Proc. of the ACM MM*. Firenze, Itália: [s.n.], 2010. p. 655–658.
- AVIDAN, S.; SHAMIR, A. Seam carving for content-aware image resizing. *ACM Trans. on Graphics*, v. 26, p. 10, 2007.
- AYRES, M. *Saiba como surgiram as câmeras fotográficas digitais*. 2007. Último acesso: [30/11/2011]. Disponível em: <<http://tecnologia.uol.com.br/produtos/ultnot/2007/08/29/ult2880u406.jhtm>>.
- BALINKSY, H. Y.; HOWES, J. R.; WILEY, A. J. Aesthetically-driven layout engine. In: *Proc. of ACM Symposium on Document Engineering*. New York, EUA: [s.n.], 2009. p. 42–51.
- BANERJEE, S.; EVANS, B. L. Unsupervised automation of photographic composition rules in digital still cameras. In: *Proc. of IS&T SPIE Conference on Sensors, Color, Cameras, and Systems for Digital Photography*. San Jose, EUA: [s.n.], 2004. p. 364–373.

- BERDAN, R. *PHOTO Composition Articles. COMPOSITION & the ELEMENTS of VISUAL DESIGN*. 2004. Último acesso: [20/11/2011]. Disponível em: <http://photoinf.com/General/Robert_Berdan/Composition_and_the_Elements_of_Visual_Design.htm>.
- BERGMANN, F.; MANSSOUR, I.; SILVEIRA, M.; OLIVEIRA, J. Automatic layout generation for digital photo albums: A user study. *Human-Computer Interaction. Users and Contexts of Use*, v. 8006, p. 117–126, 2013.
- BHATTACHARYA, S.; SUKTHANKAR, R.; SHAH, M. A framework for photo-quality assessment and enhancement based on visual aesthetics. In: *Proc. of the ACM MM 2010*. Firenze, Italy: [s.n.], 2010. p. 271–280.
- BISHOP, C. M. *Neural Networks for Pattern Recognition*. [S.l.]: Oxford University Press, 1995.
- BISHOP, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. [S.l.]: Springer-Verlag New York, Inc., 2006.
- BISSACCO, A.; YANG, M.; SOATTO, S. Fast human pose estimation using appearance and motion via multi-dimensional boosting regression. In: *Proc. of IEEE CVPR*. Minneapolis, EUA: [s.n.], 2007. p. 1–8.
- BOUTELL, M.; LUO, J.; GRAY, R. T. Sunset scene classification using simulated image recomposition. In: *Proc. of IEEE ICME*. Baltimore, EUA: [s.n.], 2003. p. 37–40.
- BRADSKI, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, v. 1, p. 122–125, 2000.
- BRUNEAU, P.; PICAROUGNE, F.; GELGON, M. Interactive unsupervised classification and visualization for browsing an image collection. *Pattern Recognition*, v. 43, p. 485–493, 2010.
- BUSSELLE, M. *Better Picture Guide to Photographing People*. [S.l.]: RotoVision, 1999.
- BYERS, Z.; DIXON, M.; SMART, W.; GRIMM, C. Say cheese!: Experiences with a robot photographer. *AAAI Magazine*, v. 25, p. 37–46, 2004.
- CARDOSO, F. H. *Um Método Automático para o Ajuste de Segmentadores de Imagens Baseados em Informação de Textura e Cor*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2009.
- CARDOSO, F. H.; GOMES, H. M. A probabilistic approach to skin detection. In: *Anais do Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens*. Belo Horizonte, Brasil: [s.n.], 2007. p. 1–2.
- CARTER, R. L. *DigiCamHistory*. 2001. Último acesso: [30/11/2011]. Disponível em: <<http://www.digicamhistory.com/>>.

- CAVALCANTI, C.; GOMES, H.; VELOSO, L.; CARVALHO, J.; JR, O. L. Automatic single person composition analysis. In: *Proc. of 18th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision - WSCG*. Plzen, Czech Republic: [s.n.], 2010. p. 229–236.
- CAVALCANTI, C. S. V. C. *Algoritmos para Composição Automática de Fotografias*. Dissertação (Mestrado) — Universidade Federal de Campina Grande, 2007.
- CAVALCANTI, C. S. V. C.; GOMES, H.; MEIRELES, R.; GUERRA, W. Towards automating photographic composition of people. In: *Proc. of IASTED Visualization, Imaging and Image Processing*. Palma de Mallorca, Espanha: [s.n.], 2006. p. 25–30.
- CAVALCANTI, C. S. V. C.; GOMES, H. M.; QUEIROZ, J. E. R. de. Combining multiple image features to guide automatic portrait cropping for rendering different aspect ratios. In: *Proc. of International Conference on Signal-Image Technology and Internet Based Systems*. Kuala Lumpur, Malásia: [s.n.], 2010. p. 66–73.
- CAVALCANTI, C. S. V. C.; GOMES, H. M.; QUEIROZ, J. E. R. de. An extended evaluation of methods for portrait cropping. *Journal of Multimedia Processing and Technologies*, v. 1, p. 245–259, 2010.
- CAVALCANTI, C. S. V. C.; QUEIROZ, J. E. R. de; GOMES, H. M. A survey on automatic techniques for enhancement and analysis of digital photography. *Journal of the Brazilian Computer Society*, v. 19, p. 341–359, 2013.
- CELIK, T.; TJAHJADI, T. Unsupervised colour image segmentation using dual-tree complex wavelet transform. *Computer Vision and Image Understanding*, v. 114, p. 813–826, 2010.
- Challenging Technologies. *dpchallenge a digital photography contest*. 2013. Último acesso: [31/01/13]. Disponível em: <<http://www.dpchallenge.com>>.
- CHANG, C.-C.; LIN, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. on Intelligent Systems and Technology*, v. 2, p. 27:1–27:27, 2011.
- CHAO, H.; TRETTER, D. R.; ZHANG, X.; ATKINS, C. B. Blocked recursive image composition with exclusion zones. In: *Proc. of ACM Symposium on Document engineering*. Manchester, Reino Unido: [s.n.], 2010. p. 111–114.
- CHEN, B. X. *Is iPhone Murdering Point-and-Shoots? Not Really*. 2011. Último acesso: [29/09/2011]. Disponível em: <<http://www.wired.com/gadgetlab/2011/04/iphone-camera/>>.
- CHEN, L.-q.; XIE, X.; FAN, X.; MA, W.-Y.; ZHANG, H.-j.; ZHOU, H.-Q. A visual attention model for adapting images on small displays. *Multimedia Systems*, v. 9, p. 353–364, 2003.
- CHEN, S.; CAO, L.; WANG, Y.; LIU, J.; TANG, X. Image segmentation by map-ml estimations. *IEEE Trans. on Image Processing*, v. 19, p. 2254–64, 2010.

- CHERRY, K. *Color Psychology: How Colors Impact Moods, Feelings, and Behaviors*. 2009. Último acesso: [10/01/2012]. Disponível em: <<http://psychology.about.com/od/sensationandperception/a/colorpsych.htm>>.
- CHUANG, Y.-Y.; CURLESS, B.; SALESIN, D. H.; SZELISKI, R. A bayesian approach to digital matting. In: *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. Kauai, EUA: [s.n.], 2001. p. 264–271.
- CIOCCA, G.; CUSANO, C.; GASPARINI, F.; SCHETTINI, R. Self-Adaptive Image Cropping for Small Displays. In: *Proc. of IEEE International Conference on Consumer Electronics*. Las Vegas, EUA: [s.n.], 2007. p. 1–2.
- CLN. *Worldwide Revenue From Digital Still Cameras to Reach US\$3.5 Billion by 2015*. 2011. Último acesso: [29/09/2011]. Atualmente indisponível. Disponível em: <http://cln-online.org/index.php?option=com_content&view=article&id=1657:still-camera-market&catid=38:research&Itemid=100>.
- COHEN-OR, D.; SORKINE, O.; GAL, R.; LEYVAND, T.; XU, Y.-Q. Color harmonization. *ACM Trans. on Graphics (Proc. of ACM SIGGRAPH)*, v. 25, p. 624–630, 2006.
- COLOR, T. *Basic Color Schemes - Introduction to Color Theory*. 2009. Último acesso: [12/12/09]. Disponível em: <<http://www.tigercolor.com/color-lab/color-theory/color-theory-intro.htm>>.
- COMANICIU, D.; MEER, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 24, p. 603–619, 2002.
- CONNORS, M.; CONNORS, K. *MorgueFile*. 2009. Último acesso: [24/11/2011]. Disponível em: <<http://www.morguefile.com>>.
- CORBIS. *Corbis Image Gallery*. 2001–2009. Último acesso: [06/09/13]. Disponível em: <<http://www.corbis.com>>.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 1, p. 273–297, 1995.
- COUZINIE-DEVY, F.; SUN, J.; ALAHARI, K.; PONCE, J. Learning to estimate and remove non-uniform image blur. In: *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: [s.n.], 2013. p. 1075–1082.
- Creative Commons. *Creative Commons*. 2012. Último acesso: [06/09/13]. Disponível em: <<http://creativecommons.org/>>.
- CUNNINGHAM, S. J.; MASOODIAN, M. Identifying personal photo digital library features. In: *Proc. of the ACM/IEEE-CS JCDL 2007*. Vancouver, BC, Canada: [s.n.], 2007. p. 400–401.

- DALIRI, M. R.; TORRE, V. Classification of silhouettes using contour fragments. *Computer Vision and Image Understanding*, v. 113, p. 1017–1025, 2009.
- DATTA, R.; JOSHI, D.; LI, J.; WANG, J. Z. Studying aesthetics in photographic images using a computational approach. In: *Proc. of the European Conference on Computer Vision*. Graz, Austria: [s.n.], 2006. p. 288–301.
- DEHGHAN, F.; ABRISHAMI-MOGHADDAM, H. Comparison of svm and neural network classifiers in automatic detection of clustered microcalcifications in digitized mammograms. In: *Machine Learning and Cybernetics, 2008 International Conference on*. Kunming, China: [s.n.], 2008. p. 756–761.
- DESTRERO, A.; MOL, C.; ODOÑO, F.; VERRI, A. A regularized framework for feature selection in face detection and authentication. *Int. J. Comput. Vision*, v. 83, p. 164–177, 2009.
- DONDIS, D. *A sintaxe da linguagem visual*. [S.l.]: Martins Fontes, 1991.
- DRUCKER, J.; MCVARISH, E. *Graphic Design History: A Critical Guide*. [S.l.]: Pearson Prentice Hall, 2010.
- DUNLEAVY, D. *A Bird's View of History: The Digital Camera and the Ever-Changing Landscape of Photojournalism*. 2006. Último acesso: [30/11/2011]. Disponível em: <<http://digitaljournalist.org/issue0602/dunleavy.html>>.
- EBERHART, R.; SHI, Y. Comparison between genetic algorithms and particle swarm optimization. *Evolutionary Programming VII*, v. 1447, p. 611–616, 1998.
- EGOROVA, M.; SAFONOV, I. Algorithms for photo book authoring. *Pattern Recognition and Image Analysis*, v. 23, p. 57–67, 2013.
- EGOROVA, M. A.; SAFONOV, I. V. Photobook generation algorithms. *Pattern Recognit. Image Anal.*, v. 21, p. 477–480, 2011.
- ENGELKE, U.; MAEDER, A. J.; ZEPERNICK, H.-J. On confidence and response times of human observers in subjective image quality assessment. In: *Proc. of IEEE ICME*. New York, EUA: [s.n.], 2009. p. 910–913.
- ESTRADA, F. J.; JEPSON, A. D. Benchmarking image segmentation algorithms. *Int. J. Comput. Vision*, v. 85, p. 167–181, 2009.
- FAIRCHILD, M. D. *Color Appearance Models*. [S.l.]: John Wiley & Sons, Ltd, 2005.
- FASEL, I.; FORTENBERRY, B.; MOVELLAN, J. A generative framework for real time object detection and classification. *Computer Vision and Image Understanding*, v. 98, p. 182–210, 2005.
- FEDOROVSKAYA, E.; NEUSTAEDTER, C.; HAO, W. Image harmony for consumer images. In: *Proc. of IEEE ICIP*. San Diego, EUA: [s.n.], 2008. p. 121–124.
- FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, v. 59, p. 167–181, 2004.

Flickr. *Flickr Photo Sharing*. 2013. Último acesso: [31/01/13]. Disponível em: <<http://www.flickr.com/>>.

FOGGIA, P.; PERCANNELLA, G.; SANSONE, C.; VENTO, M. A graph-based algorithm for cluster detection. *Int. J. Pattern Recognition and Artificial Intelligence*, v. 22, p. 843–860, 2008.

FREE Photos Bank: Free stock photos. 2010. Último acesso: [24/11/2011]. Disponível em: <<http://www.freephotosbank.com>>.

FREEMAN, M. *Photographing People*. [S.l.]: The Ilex Press Limited, 2004.

GALLAGHER, A.; CHEN, T. Clothing cosegmentation for recognizing people. In: *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. Anchorage, EUA: [s.n.], 2008. p. 1–8.

GARAU, A.; BRUNO, N.; ARNHEIM, R. *Color Harmonies*. [S.l.]: University Of Chicago Press, 1993.

GEGENFURTNER, K. R.; KIPER, D. C. Color vision. *Annual Reviews in Neuroscience*, v. 26, p. 181–206, 2003.

GEIGEL, J.; LOUI, A. Using genetic algorithms for album page layouts. *MultiMedia, IEEE*, v. 10, p. 16–27, 2003.

GEIGEL, J.; LOUI, A.; LOUI, E. Automatic page layout using genetic algorithms for electronic albuming. In: *Working Paper, Research and Development, Eastman Kodak Company*. [S.l.: s.n.], 2001. p. 1–12.

GOMES FILHO, J. *Gestalt do objeto: sistemas de leitura visual da forma*. [S.l.]: Escrituras Editora, 2000.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. [S.l.]: Addison-Wesley Longman Publishing Co., Inc., 2001.

GORELICK, L.; BASRI, R. Shape based detection and top-down delineation using image segments. *Int. J. Comput. Vision*, v. 83, p. 211–232, 2009.

GOZALI, J. P.; KAN, M.-Y.; SUNDARAM, H. How do people organize their photos in each event and how does it affect storytelling, searching and interpretation tasks? In: *Proc. of the 12th ACM/IEEE-CS joint conference on Digital Libraries*. New York, NY, USA: [s.n.], 2012. p. 315–324.

GREENSPUN, P. *History of Photography Timeline*. 2007. Último acesso: [30/11/2011]. Disponível em: <<http://photo.net/history/timeline>>.

GREENSPUN, P. *Photo.net photography community*. 2013. Último acesso: [06/09/13]. Disponível em: <<http://photo.net>>.

GRILL, T.; SCANLON, T. *Photographic Composition*. [S.l.]: New York: Amphoto Books. Watson-Guption Publications, 1990.

- GUDMUNDSSON, B. *Stockvault: we love to share photos*. 2011. Último acesso: [24/11/2011]. Disponível em: <<http://www.stockvault.net>>.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: An update. *SIGKDD Explorations Newsletter*, v. 11, p. 10–18, 2009.
- HAN, J.; AWAD, G.; SUTHERLAND, A. Automatic skin segmentation and tracking in sign language recognition. *IET-Computer Vision*, v. 3, p. 24–35, 2009.
- HASSAN, R.; COHANIM, B.; WECK, O. D.; VENTER, G. A Comparison Of Particle Swarm Optimization And The Genetic Algorithm. In: *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. Austin, EUA: [s.n.], 2005. p. 1–13.
- HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. [S.l.]: Prentice Hall, 1999.
- HEDGECOE, J. *The new manual of photography*. [S.l.]: D.K. - Dorling Kindersley Limited, 2003.
- HEDGECOE, J. *The Book of Photography*. [S.l.]: D.K. - Dorling Kindersley Limited, 2005.
- HEDGECOE, J. *New Manual of Photography*. [S.l.]: Dorling Kindersley, 2009.
- HESTER, D. *PIXEL Perfect Digital - Free Stock Photography*. 2009. Último acesso: [24/11/2011]. Disponível em: <<http://www.pixelperfectdigital.com>>.
- HOLLAND, J. H. *Adaptation in natural and artificial systems: An Introductory Analysis with Applications to biology, control and artificial intelligence*. [S.l.]: MIT Press, 1992.
- HOOK, C. *Perceptual Limitations*. 2004. Último acesso: [12/12/09]. Disponível em: <<http://everything2.com/title/perceptual+limitations>>.
- HOU, X.; ZHANG, L. Color conceptualization. In: *Proc. of the ACM MM*. Augsburg, Alemanha: [s.n.], 2007. p. 265–268.
- HUA, S.; LI, X.; ZHONG, Q. Similarity criterion for image resizing. *EURASIP Journal on Advances in Signal Processing*, v. 2011, p. 1–8, 2011.
- HURTER, B. *The Portrait Photographer's Guide to Posing*. [S.l.]: Amherst Media, 2004.
- HUYNH, D. F.; DRUCKER, S. M.; BAUDISCH, P.; WONG, C. Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. In: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: [s.n.], 2005. p. 1937–1940.
- INC., A. S. *What's new in Photoshop CS6*. 2010. Último acesso: [31/01/13]. Disponível em: <<http://www.adobe.com/products/photoshop.html>>.

- INC., A. S. *Adobe Kuler Color Wheel*. 2013. Último acesso: [05/06/2013]. Disponível em: <<https://kuler.adobe.com/>>.
- INC., G. *Google App Engine*. 2013. Último acesso: [05/06/2013]. Disponível em: <<https://developers.google.com/appengine/>>.
- ITTEN, J. *The Art of Color*. [S.l.]: Van Nostrand Reinhold Company, 1960.
- ITTEN, J. *Kunst der Farbe*. [S.l.]: Reinhold Pub. Corp., 1961.
- ITTI, L.; REES, G.; TSOTSOS, J. K. *Neurobiology of Attention*. [S.l.]: Elsevier Science Inc., 2005.
- JAIN, R. *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modeling*. [S.l.]: Wiley- Interscience, 1991.
- JI, H.; LIU, C. Motion blur identification from image gradients. In: *Proc. of the IEEE CVPR 2007*. Los Alamitos, CA, USA: [s.n.], 2008. p. 1–8.
- JONES, K. O. Comparison of genetic algorithm and particle swarm optimization. In: *Proc. of International Conference on Computer Systems and Technologies - CompSys-Tech'2005*. Varna, Bulgária: [s.n.], 2005. p. IIIA1–6.
- JR., C. A. S. L.; LAPA, C. M. F.; PEREIRA, C. M. do N.; CUNHA, J. J. da; ALVIM, A. C. M. Comparison of computational performance of ga and pso optimization techniques when designing similar systems. *Annals of Nuclear Energy*, v. 38, p. 1339 – 1346, 2011.
- KACHITVICHYANUKUL, V. On Comparison of Three Evolutionary Algorithms: GA, PSO and DE. *Industrial Engineering and Management Systems*, v. 11, p. 215–223, 2012.
- KANG, H.; LEE, S.; CHUI, C. K. Coherent line drawing. In: *ACM Symposium on Non-Photorealistic Animation and Rendering (NPAR)*. San Diego, EUA: [s.n.], 2007. p. 43–50.
- KE, Y.; TANG, X.; JING, F. The design of high-level features for photo quality assessment. In: *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. New York, EUA: [s.n.], 2006. p. 419–426.
- KENNEDY, J.; EBERHART, R. Particle swarm optimization. In: *Neural Networks, 1995. Proc., IEEE International Conference on*. Perth, Austrália: [s.n.], 1995. p. 1942–1948 vol.4.
- KOBAYASHI, S. *A Book of Colors*. [S.l.]: Kodansha Ltd., 1987.
- KOBAYASHI, S. *Color Image Scale*. [S.l.]: Kodansha Ltd., 1990.
- KOFFKA, K. *Principles of Gestalt Psychology*. [S.l.]: Routledge and Kegan Paul Ltd., 1955.

- KOZA, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection (Complex Adaptive Systems)*. [S.l.]: A Bradford Book, 1992.
- KRÖNER, A. The designcomposer: Context-based automated layout for the internet. In: *Proc. of AAAI National Conference*. Orlando, EUA: [s.n.], 1999. p. 1–5.
- KRUPPA, H.; BAUER, M. A.; SCHIELE, B. Skin patch detection in real-world images. In: *Proc. of Annual Symposium for Pattern Recognition of the DAGM*. Zurich, Suíça: [s.n.], 2002. p. 109–117.
- KUHNA, M.; KIVELÄ, I.-M.; OITTINEN, P. Semi-automated magazine layout using content-based image features. In: *Proc. of the 20th ACM MM*. New York, NY, USA: [s.n.], 2012. p. 379–388.
- LEE, S.; KIM, K.; KIM, J.-Y.; KIM, M.; YOO, H.-J. Familiarity based unified visual attention model for fast and robust object recognition. *Pattern Recogn.*, v. 43, p. 1116–1128, 2010.
- LEITE, B. de B.; PEREIRA, E. T.; GOMES, H. M.; VELOSO, L. R.; SANTOS, C. E. do N.; CARVALHO, J. M. de. A learning-based eye detector coupled with eye candidate filtering and pca features. In: *Proc. of Brazilian Symposium on Computer Graphics and Image Processing*. Belo Horizonte, Brasil: [s.n.], 2007. p. 187–194.
- LEVIN, A.; WEISS, Y. Learning to combine bottom-up and top-down segmentation. *Int. J. Comput. Vision*, v. 81, p. 105–118, 2009.
- LI, Z.; ZHOU, X.; HUANG, T. S. Spatial gaussian mixture model for gender recognition. In: *Proc. of IEEE ICIP*. Cairo, Egypt: [s.n.], 2009. p. 45–48.
- LIDWELL, W.; HOLDEN, K.; BUTLER, J. *Universal Principles of Design*. [S.l.]: Rockport Publishers, Inc., 2003.
- LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology*, v. 22, p. 1–55, 1932.
- LINDEN, R. *Algoritmos Genéticos*. [S.l.]: Brasport, 2008.
- LIU, L.; CHEN, R.; WOLF, L.; COHEN-OR, D. Optimizing photo composition. In: *Proc. of Eurographics*. Norrköping, Suécia: [s.n.], 2010. p. 469–478.
- LIU, R.; LI, Z.; JIA, J. Image partial blur detection and classification. In: *Proc. of the IEEE CVPR 2007*. Los Alamitos, CA, USA: [s.n.], 2008. p. 1–8.
- LIU, R.; LI, Z.; JIA, J. Image partial blur detection and classification. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. Anchorage, USA: [s.n.], 2008. p. 1–8.
- LOK, S.; KAN, M.-Y. Employing natural language summarization and automated layout for effective presentation and navigation of information retrieval results. In: *Proc. of International World Wide Web Conference*. Budapest, Hungria: [s.n.], 2003. p. 33–36.

- LOUI, A.; WOOD, M.; SCALISE, A.; BIRKELUND, J. Multidimensional image value assessment and rating for automated albuming and retrieval. In: *Proc. of IEEE ICIP*. San Diego, EUA: [s.n.], 2008. p. 97–100.
- LUO, W.; WANG, X.; TANG, X. Content-based photo quality assessment. In: *Proc. of the IEEE ICCV 2011*. Barcelona, Espanha: [s.n.], 2011. p. 2206–2213.
- LUO, Y.; TANG, X. Photo and video quality evaluation: Focusing on the subject. In: *Proc. of European Conference on Computer Vision*. Marseille, França: [s.n.], 2008. p. 386–399.
- LUX, M.; KOGLER, M.; FABRO, M. del. Why did you take this photo: a study on user intentions in digital photo productions. In: *Proc. of the ACM SAPMIA 2010*. Firenze, Italy: [s.n.], 2010. p. 41–44.
- MAIK, V.; PAIK, D.; LIM, J.; PARK, K.; PAIK, J. Hierarchical pose classification based on human physiology for behaviour analysis. *IET-Computer Vision*, v. 4, p. 12–24, 2010.
- MANTIUK, R. K.; TOMASZEWSKA, A.; MANTIUK, R. Comparison of four subjective methods for image quality assessment. *Comput. Graphics Forum*, v. 31, p. 2478–2491, 2012.
- MATSUDA, Y. *Color Design*. [S.l.]: Asakura Shoten, 1995.
- MILLER, A. D.; EDWARDS, W. K. Give and take: a study of consumer photo-sharing culture and practice. In: *Proc. of the ACM SIGCHI 2007*. San Jose, California, USA: [s.n.], 2007. p. 347–356.
- MORENO, L. *O design equilibrado. O contraste*. 2007. Último acesso: [30/11/2011]. Disponível em: <<http://www.criarweb.com/artigos/design-equilibrado-contraste.html>>.
- MORENO, L. *O design equilibrado. Os alinhamentos*. 2007. Último acesso: [30/11/2011]. Disponível em: <<http://www.criarweb.com/artigos/design-equilibrado-alinhamentos.html>>.
- MORTON, J. *Colors That Sell: Tried and Tested Color Schemes*. [S.l.]: Colorcom, 2004.
- MORTON, J. *Color Matters*. 2008. Último acesso: [10/12/09]. Disponível em: <<http://www.colormatters.com/colortheory.html>>.
- MORTON, J. *A Guide to Color Symbolism*. [S.l.]: Colorcom, 2008.
- MU, Y.; YAN, S.; LIU, Y.; HUANG, T.; ZHOU, B. Discriminative local binary patterns for human detection in personal album. In: *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*. Anchorage, EUA: [s.n.], 2008. p. 1–8.
- MUCIENTES, M.; BUGARÍN, A. People detection through quantified fuzzy temporal rules. *Pattern Recogn.*, v. 43, p. 1441–1453, 2010.

- MUNSELL, A. H. *A Color Notation*. [S.l.]: G. H. Ellis Co., 1905.
- MUNSELL, A. H. *A Grammar of Colors*. [S.l.]: Van Nostrand Reinhold Company, 1969.
- NAAMAN, M.; HARADA, S.; WANG, Q.; GARCIA-MOLINA, H.; PAEPCKE, A. Context data in geo-referenced digital photo collections. In: *Proceedings of the 12th annual ACM international conference on Multimedia*. New York, NY, USA: [s.n.], 2004. p. 196–203.
- NEUMANN, L.; NEUMANN, A. Color style transfer techniques using hue, lightness and saturation histogram matching. In: *Proc. of The International Symposium on Computational Aesthetics in Graphics, Visualization and Imaging*. Girona, Espanha: [s.n.], 2005. p. 111–122.
- OBRADOR, P. Automatic color scheme picker for document templates based on image analysis and dual problem. In: *Proc. of SPIE Conference on Digital Publishing*. San Jose, EUA: [s.n.], 2006. p. 64–73.
- OBRADOR, P. Region based image appeal metric for consumer photos. In: *Proc. of the IEEE Workshop on Multimedia Signal 2008*. Queensland, Australia: [s.n.], 2008. p. 696–701.
- OBRADOR, P.; MORONEY, N. Automatic image selection by means of a hierarchical scalable collection representation. In: *Proc. SPIE 7257 Visual Communications and Image Processing*. San Diego, EUA: [s.n.], 2009. p. 72570W–72570W–12.
- OKAZAKI, R. Y. *Creating a Storyboard for Video Production*. 1998. Último acesso: [03/06/07]. Disponível em: <<http://www2.hawaii.edu/~ricky/etec/basicshot.html>>.
- OLIVEIRA, C. J. S.; ARAÚJO, A. D. A.; JR., C. A. S.; GOMES, D. R. Classifying images collected on the world wide web. In: *Proc. of Brazilian Symposium on Computer Graphics and Image Processing*. Fortaleza, Brasil: [s.n.], 2002. p. 327–334.
- ORBANZ, P.; BUHMANN, J. M. Nonparametric bayesian image segmentation. *Int. J. Comput. Vision*, v. 77, p. 25–45, 2008.
- OWEN, D.; OWEN, R.; HOLT, A. *Convert 4x3 Video to 16x9 Widescreen*. 2008. Último acesso: [20/01/2012]. Disponível em: <<http://www.mediacollege.com/video/aspect-ratio/4x3-to-16x9.html>>.
- P., A. *Understanding Aspect Ratio*. 2010. Último acesso: [20/01/2012]. Disponível em: <http://www.crutchfield.com/S-Rf3os1wabSq/learn/learningcenter/home/aspect_ratio.html>.
- PAISITKRIANGKRAI, S.; SHEN, C.; ZHANG, J. Performance evaluation of local features in human classification and detection. *IET-Computer Vision*, v. 2, p. 236–246, 2008.

PANDA, S.; PADHY, N. P. Comparison of particle swarm optimization and genetic algorithm for facts-based controller design. *Appl. Soft Comput.*, v. 8, p. 1418–1427, 2008.

PANTOFARU, C.; HEBERT, M. *A Comparison of Image Segmentation Algorithms*. Pittsburgh, PA, 2005.

PEREIRA, E. T.; GOMES, H. M. Guiding a bottom-up visual attention mechanism to locate specific image regions using a distributed genetic optimization. In: *Proc. of Iberoamerican Congress on Pattern Recognition*. Cancun, Mexico: [s.n.], 2006. p. 257–266.

PEREIRA, E. T.; GOMES, H. M.; FLORENTINO, V. F. C. Bottom-up visual attention guided by genetic algorithm optimization. In: *Proc. of IASTED International Conference on Signal and Image Processing*. Honolulu, EUA: [s.n.], 2006. p. 228–233.

PETERSON, E. *Logo Critiques*. 2009. Último acesso: [10/01/2012]. Disponível em: <http://www.logocritiques.com/resources/color_psychology_in_logodesign>.

PICCOLI, R. F. B. *Um Algoritmo Determinístico para Adaptação Automática de Documentos de Conteúdo Variável*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio Grande do Sul, 2010.

PIERRARD, J. sébastien; VETTER, T. Skin detail analysis for face recognition. In: *Proc. of IEEE CVPR 2007*. Minneapolis, EUA: [s.n.], 2007. p. 1–8.

PRATT, W. K. *Digital Image Processing*. [S.l.]: Wiley-Interscience, 2007.

PRO, C. W. *Color Meaning*. 2008. Último acesso: [10/01/12]. Disponível em: <<http://www.color-wheel-pro.com/color-meaning.html>>.

PURVIS, L. A genetic algorithm approach to automated custom document assembly. In: *Proc. of international Workshop on Intelligent Systems Design and application*. Atlanta, Georgia: [s.n.], 2002. p. 131–136.

QIN, A. K.; CLAUSI, D. A. Multivariate image segmentation using semantic region growing with adaptive edge penalty. *IEEE Trans. on Image Processing*, v. 19, p. 2157–70, 2010.

QUINLAN, J. R. Induction of decision trees. *Machine Learning*, v. 1, p. 81–106, 1986.

RABBATH, M.; SANDHAUS, P.; BOLL, S. Automatic creation of photo books from stories in social media. In: *Proc. of ACM SIGMM workshop on Social media*. New York, NY, USA: [s.n.], 2010. p. 15–20.

RABBATH, M.; SANDHAUS, P.; BOLL, S. Automatic creation of photo books from stories in social media. *ACM Trans. on Multimedia, Computing, Communications and Applications*, v. 7S, p. 27:1–27:18, 2011.

RESEARCH, P. M. *PMA Camera/Camcorder and Digital Imaging Survey*. [S.l.], 2008.

- ROCKEWELL, K. *Nikon DSLR History*. 2011. Último acesso: [29/09/2011]. Disponível em: <<http://www.kenrockwell.com/nikon/dslr.htm>>.
- RODDEN, K.; WOOD, K. R. How do people manage their digital photographs? In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: [s.n.], 2003. p. 409–416.
- ROUSSON, M.; PARAGIOS, N. Prior knowledge, level set representations & visual grouping. *Int. J. Comput. Vision*, v. 76, p. 231–243, 2008.
- ROWLEY, H. A.; BALUJA, S.; KANADE, T. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 20, p. 23–28, 1998.
- ROWLEY, H. A.; BALUJA, S.; KANADE, T. Rotation invariant neural network-based face detection. *Computer Vision and Pattern Recognition*, v. 20, p. 38–44, 1998.
- RUBNER, Y.; TOMASI, C.; GUIBAS, L. J. The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision*, v. 40, p. 2000, 2000.
- RYOO, M. S.; AGGARWAL, J. K. Semantic representation and recognition of continued and recursive human activities. *Int. J. Comput. Vision*, v. 82, p. 1–24, 2009.
- SANDHAUS, P.; RABBATH, M.; BOLL, S. Employing aesthetic principles for automatic photo book layout. In: *Proc. of the 17th international conference on Advances in multimedia modeling - Volume Part I*. Berlin, Heidelberg: [s.n.], 2011. p. 84–95.
- SANTELLA, A.; AGRAWALA, M.; DECARLO, D.; SALESIN, D.; COHEN, M. Gaze-based interaction for semi-automatic photo cropping. In: *Proc. of the ACM SIGCHI Conference on Human Factors in Computing Systems*. Montréal, Canadá: [s.n.], 2006. p. 771–780.
- SAUVAGET, C.; VITTAUT, J.-N.; SUAREZ, J.; BOYER, V.; MANUEL, S. Automated colorization of segmented images based on color harmony. *JMPT*, v. 1, p. 228–244, 2010.
- SAVAKIS, A. E.; ETZ, S. P.; LOUI, A. C. Evaluation of image appeal in consumer photography. In: *Proc. of the SPIE Human Vision and Electronic Imaging V*. San Jose, EUA: [s.n.], 2000. p. 111–120.
- SAWANT, N.; MITRA, N. J. Color harmonization for videos. *Computer Vision, Graphics & Image Processing, Indian Conference on*, v. 1, p. 576–582, 2008.
- SERRANO, N.; SAVAKIS, A.; LUO, J. A computationally efficient approach to indoor/outdoor scene classification. In: *Proc. of the ICPR 2002*. Washington, EUA: [s.n.], 2002. p. 146–149.
- SETLUR, V.; TAKAGI, S.; RASKAR, R.; GLEICHER, M.; GOOCH, B. Automatic image retargeting. In: *MUM '05: Proc. of the 4th international conference on Mobile and ubiquitous multimedia*. Christchurch, Nova Zelândia: [s.n.], 2005. p. 59–68.

- SHEN, C. T.; LIU, J. C.; SHIH, S. W.; HONG, J. S. Towards intelligent photo composition-automatic detection of unintentional dissection lines in environmental portrait photos. *Expert Systems with Applications*, v. 36, p. 9024–9030, 2009.
- SHLENS, J. *A Tutorial on Principal Component Analysis*. 2005. Último acesso: [06/09/2013]. Disponível em: <<http://www.snl.salk.edu/~shlens/pca.pdf>>.
- SRIDHARAN, H.; SUNDARAM, H.; BRUNGART, J. *Automated Design of Paper Workbooks, for Electronic Learning Environments*. [S.l.], 2004.
- STEIN, A.; STEPLETON, T.; HEBERT, M. Towards unsupervised whole-object segmentation: Combining automated matting with boundary detection. In: *Proc. of IEEE CVPR 2008*. Anchorage, EUA: [s.n.], 2008. p. 1–8.
- SUH, B.; LING, H.; BEDERSON, B. B.; JACOBS, D. W. Automatic thumbnail cropping and its effectiveness. In: *Proc. of the ACM Symposium on User Interface Software and Technology*. Vancouver, Canadá: [s.n.], 2003. p. 95–104.
- SULLIVAN, T.; MONSON, K. *Why Aren't Camera Companies Making Money?* 2005. Último acesso: [29/09/2011]. Disponível em: <<http://www.pcmag.com/article2/0,2817,1885729,00.asp#fbid=bO5iykiU7MG>>.
- TILLEY, A. R. *The Measure of Man and Woman: Human Factors in Design*. [S.l.]: Wiley, 2001.
- TOKUMARU, M.; MURANAKA, N.; IMANISHI, S. Color design support system considering color harmony. In: *IEEE International Conference on Fuzzy Systems*. Honolulu, EUA: [s.n.], 2002. p. 378–383.
- TONG, H.; LI, M.; ZHANG, H.; ZHANG, C. Blur detection for digital images using wavelet transform. In: *Proc. of IEEE ICME*. Taipei, Taiwan: [s.n.], 2004. p. 17–20.
- TONG, H.; LI, M.; ZHANG, H.-J.; HE, J.; ZHANG, C. Classification of digital photos taken by photographers for home users. In: *In Proc. of Pacific Rim Conference on Multimedia*. Tokyo, Japão: [s.n.], 2004. p. 198–205.
- TREMBLAY, P. *Basic rules of photography*. 2003. Último acesso: [03/06/07]. Disponível em: <<http://trem.ca/learning.html>>.
- TSAO, W.-K.; LEE, A. J. T.; LIU, Y.-H.; CHANG, T.-W.; LIN, H.-H. A data mining approach to face detection. *Pattern Recogn.*, v. 43, p. 1039–1049, 2010.
- Turbo photo inc. *Free Stock Images*. 2011. Último acesso: [05/09/2013]. Disponível em: <<http://www.turbophoto.com/FreeStockImages>>.
- VAPNIK, V. N. *The Nature of Statistical Learning Theory*. [S.l.]: Springer, 1999.
- VAQUERO, D.; TURK, M.; PULLI, K.; TICO, M.; GELFAND, N. A survey of image retargeting techniques. In: *Proc. of SPIE Applications of Digital Image Processing*. San Diego, EUA: [s.n.], 2010. p. 1–15.

- VEZHNEVETS, V.; SAZONOV, V.; ANDREEVA, A. A survey on pixel-based skin color detection techniques. In: *Proc. of GRAPHICON International Conference on Computer Graphics and Vision*. Moscow, Rússia: [s.n.], 2003. p. 85–92.
- VIOLA, P.; JONES, M. Rapid object detection using a boosted cascade of simple features. In: *Proc. of the IEEE CVPR 2001*. Kauai, EUA: [s.n.], 2001. p. 511–518.
- WALL, M. *GAlib Genetic Algorithm Library*. 2009. Último acesso: [31/01/13]. Disponível em: <<http://lancet.mit.edu/ga/>>.
- WALLICK, M. N. *AUTOMATIC ORGANIZATION OF LARGE PHOTO COLLECTIONS*. Tese (Doutorado) — UNIVERSITY OF WISCONSIN - MADISON, 2007.
- WALPOLE, R. E.; MYERS, R. H.; MYERS, S. L.; YE, K. *Probability & Statistics for Engineers & Scientists, 8 ed.* [S.l.]: Prentice Hall., 2008.
- WANG, C.; ZHANG, R.; DENG, F. Image composition with color harmonization. *Chin. Opt. Lett.*, v. 7, p. 483–485, 2009.
- WANG, H.; OLIENSIS, J. Generalizing edge detection to contour detection for image segmentation. *Computer Vision and Image Understanding*, v. 114, p. 731–744, 2010.
- WANG, J.; JIA, Y.; HUA, X.-S.; ZHANG, C.; QUAN, L. Normalized tree partitioning for image segmentation. In: *Proc. of the IEEE CVPR 2008*. Anchorage, EUA: [s.n.], 2008. p. 1–8.
- WANG, P.; JI, Q. Multi-view face and eye detection using discriminant features. *Computer Vision and Image Understanding*, v. 105, p. 99–111, 2007.
- WANG, X.-F.; HUANG, D.-S.; XU, H. An efficient local chan-veye model for image segmentation. *Pattern Recogn.*, v. 43, p. 603–618, 2010.
- WILLIAMS, R. *Design Para Quem Não É Designer*. [S.l.]: Ed. Callis, 2003.
- WU, P.; O'BRIEN-STRAIN, E.; MESSENLEHNER, B.; TRETTER, D. Photobook creation and social sharing on facebook. In: *Proc. of the IEEE ICME 2011*. Washington, DC, USA: [s.n.], 2011. p. 1–4.
- XIAO, J.; LYONS, N.; ATKINS, C. B.; GAO, Y.; CHAO, H.; ZHANG, X. iphotobook: creating photo books on mobile devices. In: *Proc. of the ACM MM*. Firenze, Itália: [s.n.], 2010. p. 1551–1554.
- XIAO, J.; ZHANG, X.; CHEATLE, P.; GAO, Y.; ATKINS, C. B. Mixed-initiative photo collage authoring. In: *Proceeding of the 16th ACM MM*. Vancouver, Canada: [s.n.], 2008. p. 509–518.
- XUE, S.-F.; TANG, H.; TRETTER, D.; LIN, Q.; ALLEBACH, J. Automatic photobook: focusing on image selection and image layout based on content and composition. In: *Proc. SPIE 8664*. San Diego, EUA: [s.n.], 2013. p. 86640D–86640D–7.
- YAN, S.; WANG, H.; LIU, J.; TANG, X.; HUANG, T. S. Misalignment-robust face recognition. *IEEE Trans. on Image Processing*, v. 19, p. 1087–1096, 2010.

- YANG, Y.; XU, D.; NIE, F.; YAN, S.; ZHUANG, Y. Image clustering using local discriminant models and global integration. *IEEE Trans. on Image Processing*, v. 19, p. 2761–2773, 2010.
- YANULEVSKAYA, V.; GEMERT, J. van; ROTH, K.; HERBOLD, A.; SEBE, N.; GEUSEBROEK, J. Emotional valence categorization using holistic image features. In: *Proc. of IEEE ICIP*. San Diego, EUA: [s.n.], 2008. p. 101–104.
- YAO, L.; SURYANARAYAN, P.; QIAO, M.; WANG, J.; LI, J. OSCAR: On-Site composition and aesthetics feedback through exemplars for photographers. *Int. J. Comput. Vision*, v. 96, p. 1–31, 2011.
- YING, Z.; GUANGYAO, L.; XIEHUA, S.; XINMIN, Z. Geometric active contours without re-initialization for image segmentation. *Pattern Recogn.*, v. 42, p. 1970–1976, 2009.
- YU, Z.; AU, O. C.; ZOU, R.; YU, W.; TIAN, J. An adaptive unsupervised approach toward pixel clustering and color image segmentation. *Pattern Recogn.*, v. 43, p. 1889–1906, 2010.
- ZENG, G.; GOOL, L. V. Multi-label image segmentation via point-wise repetition. In: *Proc. of the IEEE CVPR 2008*. Anchorage, EUA: [s.n.], 2008. p. 1–8.
- ZHANG, H.; FRITTS, J. E.; GOLDMAN, S. A. Image segmentation evaluation: A survey of unsupervised methods. *Comput. Vis. Image Underst.*, v. 110, p. 260–280, 2008.
- ZHANG, M.; ZHANG, L.; SUN, Y.; FENG, L.; MA, W. Auto Cropping for Digital Photographs. In: *Proc. of the IEEE ICME 2005*. Amsterdã, Holanda: [s.n.], 2005. p. 438–441.
- ZHAO, W.-L.; NGO, C.-W. Flip-invariant sift for copy and object detection. *IEEE Trans. on Image Processing*, v. 22, p. 980–991, 2013.
- ZHAO, W.-L.; WU, X.; NGO, C.-W. On the annotation of web videos by efficient near-duplicate search. *IEEE Trans. on Multimedia*, v. 12, p. 448–461, 2010.

Apêndice A

Decisões de projeto

Ao longo deste documento, foram descritas diversas decisões tomadas durante o desenvolvimento desta pesquisa. De forma a deixar o texto principal mais objetivo e fluente, as decisões e explicações indiretamente relacionadas ao assunto foram reposicionadas neste apêndice.

Na Seção A.1, são apresentados os critérios adotados para a seleção do algoritmo de classificação de imagens com relação ao espaço negativo. Na Seção A.2, são apresentados os argumentos pela utilização dos algoritmos genéticos como o algoritmo de otimização utilizado em toda a pesquisa. Na Seção A.3, apresentam-se os motivos pela escolha da plataforma *Web GWT* para o desenvolvimento das interfaces utilizadas nos experimentos. Por fim, na Seção A.4, são apresentadas as considerações finais deste apêndice.

A.1 Escolha do algoritmo de classificação para o método do espaço negativo

Buscou-se uma abordagem que permitisse a generalização a partir da aprendizagem das diversas amostras de histograma, referentes às classes de fotografias consideradas. De modo a construir a base de treinamento necessária para a aprendizagem da rede neural, foi selecionado um conjunto de fotografias de treinamento, contendo as 450 fotografias, descritas anteriormente. O conjunto foi dividido em 338 fotografias para o treinamento

e 112 para os testes, sendo feita uma divisão igualitária entre as classes (aquelas que estão e aquelas que não estão em conformidade com a regra do Espaço Negativo).

Com a utilização da abordagem proposta, a qual utiliza RNA, foi realizado o treinamento com as 338 fotografias, por intermédio da ferramenta Weka (HALL et al., 2009). Algoritmos genéticos foram utilizados na definição da arquitetura da rede. Com o intuito de minimizar erros decorrentes da inicialização aleatória dos pesos das RNA, o experimento foi repetido 9 vezes, totalizando 10 treinamentos. Foi obtida uma taxa média de acerto de 75,08% com desvio padrão de 3,40. Com um intervalo de confiança de 2,11 e tendo-se 95% de confiança, a média deste treinamento variou entre 72,97% e 77,19%.

Adicionalmente, decidiu-se utilizar outro método, também muito utilizado em problemas similares: máquinas de vetor de suporte (*Support Vector Machines - SVM* (VAPNIK, 1999)). O SVM foi escolhido em decorrência da popularidade, da facilidade de utilização e, principalmente, da qualidade da classificação em problemas que não requeiram distribuições específicas, a exemplo dos problemas comumente encontrados em áreas conexas a esta pesquisa (DATTA et al., 2006; ROWLEY; BALUJA; KANADE, 1998b; DEHGHAN; ABRISHAMI-MOGHADDAM, 2008).

Em virtude do número de entradas na Rede Neural (em que cada entrada corresponde a um *bin* do histograma), foi testado um algoritmo destinado à redução da quantidade de entradas por intermédio da análise dos componentes principais (PCA). Após a realização da análise, foram escolhidas 10 entradas das 100 iniciais, aquelas que correspondiam a 96% da variância dos dados. Posteriormente, foi realizado um novo treinamento, desta vez com os dados resultantes do PCA. Novamente, o treinamento foi executado por 10 vezes, resultando em um acerto médio de 80,625%, com um desvio padrão de 1,78 e intervalo de confiança de 1,10 tendo-se 95% de confiança. Desta forma, o resultado do treinamento variou entre 79,51% e 81,73%, confirmando a superioridade da utilização do PCA.

Tendo-se optado pela utilização de algoritmos inteligentes para a classificação das fotografias, existem três fatores de projeto que precisaram ser avaliados em maior detalhe:

1. **Fator A.** O método de aprendizagem a ser utilizado, RNA (HAYKIN, 1999) ou

Máquinas de Vetor de Suporte (SVM) (VAPNIK, 1999), uma vez que ambos os métodos têm sido largamente utilizados na literatura e podem produzir resultados diferentes, dependendo do tipo de informação;

2. **Fator B.** A utilização ou não da Análise dos Componentes Principais (PCA) (HAYKIN, 1999), uma vez que, em havendo alta correlação entre os dados, a utilização do PCA pode trazer ganhos significativos ao método de aprendizagem; e
3. **Fator C.** A consideração da simetria dos histogramas de projeção horizontais das fotografias, pois, uma vez que existem dois lados que podem apresentar o espaço negativo, pode ser informativo para o classificador a apresentação desses dados sem a simetria.

Foi realizado um projeto experimental com 2^3 combinações, dado que envolveu 3 fatores, de modo a verificar os fatores de maior importância. Avaliou-se (1) o classificador utilizado, (2) a utilização ou não da análise dos componentes principais e (3) a consideração da simetria ou não dos histogramas de projeção horizontais, cada um com duas possibilidades de ocorrência. Os rótulos utilizados são apresentados na Tabela A.1.

Tabela A.1: Associação entre fatores e níveis aos rótulos que serão utilizados.

Rótulo	Fator A	Fator B	Fator C
-1	Utilização do PCA	RNA	Simetrização
1	Não Utilização do PCA	SVM	Não simetrização

Em resumo, foram realizados oito experimentos, variando-se os três fatores apresentados. De modo a simplificar a realização dos experimentos, foi utilizada a ferramenta Weka (HALL et al., 2009), a qual possui implementações para o cálculo do PCA, RNA e procedimentos destinadas à integração com a biblioteca libSVM (CHANG; LIN, 2011), a qual consiste em uma implementação do algoritmo SVM.

As RNA podem apresentar ligeiras variações nos resultados, de acordo com a inicialização dos seus pesos. Entretanto, a implementação usada pelo Weka inicializa os pesos aleatoriamente, de acordo com a semente fornecida ou atribuindo o valor 0 a

todos os pesos. Na Tabela A.2, são apresentados os percentuais obtidos para cada combinação no experimento.

Tabela A.2: Resultados dos experimentos dadas as combinações.

	Com PCA		Sem PCA	
	Simetrização	Não simetrização	Simetrização	Não-Simetrização
RNA	76,44%	86,22%	68,44%	74,44%
SVM	75,78%	81,78%	48,89%	48,89%

Os resultados do processo de classificação foram obtidos por meio de Validação Cruzada, realizada em 10 desdobramentos.

A partir dos fatores e dos rótulos, pôde-se produzir a tabela de sinais (JAIN, 1991), a ser utilizada no cálculo dos efeitos associados a cada fator, sendo que o efeito o percentual que um fator isoladamente explica do resultado com todos os fatores combinados. A Tabela de Sinais, apresentada na Tabela A.3, foi construída a partir dos valores obtidos.

Tabela A.3: Tabela de Sinais para o problema. A, B e C indicam os fatores e as possíveis combinações. Y o percentual de acerto para cada uma das combinações e I o sinal para a soma de todos os efeitos.

I	A	B	C	AB	AC	BC	ABC	Y
1	-1	-1	-1	1	1	1	-1	76,44
1	1	-1	-1	-1	-1	1	1	68,44
1	-1	1	-1	-1	1	-1	1	75,78
1	1	1	-1	1	-1	-1	-1	48,89
1	-1	-1	1	1	-1	-1	1	86,22
1	1	-1	1	-1	1	-1	-1	74,44
1	-1	1	1	-1	-1	1	-1	81,78
1	1	1	1	1	1	1	1	48,89
560,88	-79,56	-50,2	21,78	-40	-9,78	-9,78	-2,22	Total
70,11	-9,945	-6,275	2,7225	-5	-1,2225	-1,2225	-0,2775	$\frac{\text{Total}}{8}$

De acordo com JAIN (1991), a soma quadrática total (SST) é dada por:

$$SST = 2^k(q_A^2 + q_B^2 + q_C^2 + q_{AB}^2 + q_{AC}^2 + q_{BC}^2 + q_{ABC}^2) \quad (\text{A.1})$$

sendo q_A , q_B , q_C , q_{AB} , q_{AC} , q_{BC} e q_{ABC} as médias por fator, os quais são obtidos a

partir da Tabela de Sinais, conforme explicitado na Equação A.2.

$$SST = 2^3 \times ((-9,945)^2 + (-6,275)^2 + (-2,7225)^2 + (-5)^2 + (-1,2225)^2 + (-1,2225)^2 + (-0,2775)^2) \quad (\text{A.2})$$

Resolvendo-se a Equação A.2, obtém-se o resultado exibido na Equação A.3.

$$SST = 8 \times (98,903 + 39,376 + 7,412 + 25 + 1,494 + 1,494 + 0,077) = 1390,053 \quad (\text{A.3})$$

Os efeitos individuais de cada fator (SS_N) são calculados a partir da Equação A.4.

$$SS_N = 2^k \times \frac{q_N^2}{SST} \quad (\text{A.4})$$

Assim, os efeitos individuais de cada fator são: SS_A 791,22/1390,053 (56,92%), SS_B 315,0015/1390,053 (22,66%), SS_C 59,296/1390,053 (4,27%), SS_{AB} 200/1390,053 (14,38%), SS_{AC} 11,956/1390,053 (0,86%), SS_{BC} 11,956/1390,053 (0,86%), SS_{ABC} 0,61/1390,053 (0,04%). Com base nos dados anteriormente apresentados, pode-se perceber que a melhor média de classificação obtida é de 70,11%. Classificando os fatores por ordem decrescente de importância dos efeitos, obtém-se:

1. Utilização do PCA (56,92%);
2. Utilização do classificador usando Redes Neurais (22,66%);
3. Combinação entre a análise dos componentes principais e o classificador utilizado (14,4%);
4. Não simetriação dos dados (4,26%);
5. Combinação de PCA e simetriação dos dados (0,86%);
6. Combinação de classificadores e simetriação dos dados (0,86%);
7. Combinação dos três fatores (0,0043%).

Uma vez que Fator A (utilização ou não de PCA) foi o mais significativo no cálculo do resultado (56,92%), dado que o cálculo teve resultado negativo e, ainda que o rótulo

negativo foi atribuído à utilização do PCA, pode-se afirmar que este é o fator mais significativo para a melhoria dos resultados.

Conclusões similares podem ser formuladas no tocante aos demais fatores: o uso de RNA pode aumentar em 6,27% o resultado da classificação, o que explica 22,66% do resultado final; também pode ser encontrada alta influência na combinação entre o PCA e o classificador de até 14,4% no resultado final. Assim, dependendo da combinação utilizada, o resultado de classificação médio pode variar em até 5%. Como também era de se esperar, ambos os classificadores utilizados são capazes de “aprender” a simetria, havendo baixa influência sobre o resultado, o que explica a variação em torno de 4% dos resultados.

A.2 Decisão sobre o método de otimização utilizado

Ao longo desta pesquisa, em muitas ocasiões, foi necessária a utilização de algoritmos de otimização. Nesta seção, discutem-se os métodos destinados à otimização e à decisão sobre qual destes métodos é o mais apropriado, de acordo com os propósitos desta pesquisa.

Existem diversos métodos destinados à otimização (e.g., derivação da função objetivo e cálculo do gradiente das funções). Entretanto, em muitas situações não é simples ou não é possível realizar tais cálculos nas funções objetivos. Adicionalmente, deve-se ressaltar o número de possíveis parâmetros requeridos por tais funções, o que dificultaria na obtenção dos resultados.

Em situações similares às aquelas descritas anteriormente, foram desenvolvidos métodos estocásticos, com os quais tenta-se aproximar tais funções. Tal aproximação é comumente realizada a partir da geração aleatória de entradas e o consequente cálculo da adequação de tais valores à função objetivo originalmente proposta.

Neste campo, duas abordagens são comumente utilizadas: os AG (HOLLAND, 1992) e a PSO (KENNEDY; EBERHART, 1995). Ambas as abordagens são evolucionárias e tem como objetivo realizar a busca pela solução adequada a partir de um conjunto de valores previamente testado. Os métodos evolucionários tem como vantagem a menor susceptibilidade a resultados sub-ótimos, uma vez que estas, ao contrário

das funções de gradiente, exploram de forma mais ampla o espaço de busca e a não existência de restrições às funções objetivo.

Os AG foram propostos por Holland (1992) e tem como objetivo realizar a evolução dos indivíduos a partir do cruzamento entre os mais aptos, remoção dos menos aptos das gerações seguintes e a existência de mutações destes indivíduos. Por sua vez, no algoritmo do PSO (KENNEDY; EBERHART, 1995), a evolução representa o movimento de um enxame, no qual as partículas mais aptas ao problema movem-se mais rapidamente, induzindo o movimento das demais partículas.

Ambas as abordagens realizam a otimização pela experimentação de indivíduos gerados inicialmente de forma aleatória, exploram o espaço de busca, evitam mínimos locais e não garantem a obtenção de um resultado ótimo.

Diversos estudos comparativos têm sido realizados utilizando-se ambas as abordagens (JR. et al., 2011; EBERHART; SHI, 1998; PANDA; PADHY, 2008; HASSAN et al., 2005; JONES, 2005; KACHITVICHYANUKUL, 2012). O objetivo destes estudos é verificar se (1) ambas as abordagens conduzem a resultados equivalentes, (2) se ambas levam a resultados com qualidade similar e (3) se ambas obtém este resultado em um número equivalente de testes da função objetivo.

A conclusão global destes estudos é que ambas as abordagens conduzem a resultados equivalentes (HASSAN et al., 2005) e de alta qualidade (HASSAN et al., 2005), o que permite concluir que, em termos de resultados, AG e PSO são equivalentes.

Com relação ao terceiro teste, o número de execuções da função objetivo necessário à convergência do resultado, pode-se dividir as pesquisas relacionadas como aquelas que indicam que o PSO realiza com um menor número de execuções (HASSAN et al., 2005; JR. et al., 2011) e aquelas que obtiveram melhores resultados com os AG (PANDA; PADHY, 2008; JONES, 2005). De acordo com Hassan et al. (2005), o desempenho dos métodos é dependente do problema. Também se atestam que ambas as abordagens conduzem a resultados em um tempo aceitável (JONES, 2005).

Com base nestes resultados, decidiu-se pela utilização dos AG, uma vez que (1) os resultados são equivalentes aos produzidos pelo PSO e obtidos em tempo aceitável (HASSAN et al., 2005; JONES, 2005), (2) os AG dispõem de uma biblioteca estável, consolidada e desenvolvida na linguagem C++, a qual foi utilizada no restante desta

pesquisa (WALL, 2009) e (3) uma vez que atestou-se que o desempenho é dependente do problema (HASSAN et al., 2005), não há garantias de que o esforço de modificação do núcleo do sistema proposto nesta pesquisa produzirá resultados significativamente melhores, assim como foi atestado por Panda e Padhy (2008) e Jones (2005).

A.3 Plataforma Web para a realização de experimentos subjetivos

Ao longo desta pesquisa, foram utilizadas ferramentas para a obtenção da opinião de participantes. De modo a ter neste processo maior abrangência e velocidade, decidiu-se por utilizar plataformas Web. Nesta seção, discutem-se os motivos da escolha da tecnologia Web utilizada.

Os sistemas de suporte aos experimentos realizados nesta pesquisa foram desenvolvidos utilizando-se a tecnologia GWT (*Google Web Toolkit*¹¹). Optou-se por tal tecnologia uma vez que esta:

1. Permite o desenvolvimento de aplicações Web dinâmicas;
2. Provê hospedagem gratuita no *Google App Engine* (INC., 2013b), o qual permite armazenar dados de forma persistente e possui alta disponibilidade, o que não seria facilmente obtido utilizando-se apenas a tecnologia existente no Laboratório de Visão Computacional da UFCG, onde esta pesquisa foi desenvolvida;
3. Utiliza a linguagem Java tanto para o desenvolvimento da aplicação do lado do cliente como do lado do servidor, o que acelera o desenvolvimento de aplicações;
4. Apresenta grande quantidade de colaboradores, os quais proveem componentes para diversas finalidades e é patrocinada por uma das maiores empresas de tecnologia (Google); e
5. É gratuita.

¹¹<http://www.gwtproject.org/>

A.4 Considerações finais

Neste apêndice, foram apresentados os experimentos realizados para a decisão de parâmetros e métodos utilizados ao longo da pesquisa. Tais experimentos foram deslocados para este apêndice de modo a dar mais fluidez ao texto e, ao mesmo tempo, não omitir definitivamente tais informações. Conclui-se que, para o treinamento do método destinado à classificação do espaço negativo, a combinação de RNA e PCA resulta nos melhores resultados. Também se opta pela utilização dos AG, apesar do PSO apresentar, em alguns casos, convergência mais rápida. Desta forma, pretende-se manter a consistência na utilização do método e simplicidade de adaptação da biblioteca utilizada nos diversos problemas de otimização necessários no decorrer da pesquisa. Por fim, verificou-se que a plataforma *Web GWT* serve aos propósitos desta pesquisa, no tocante à produção rápida de interfaces de votação.

Apêndice B

Algoritmo evolucionário para a obtenção de métodos de processamento de imagens

Neste apêndice, é descrito um algoritmo baseado em Computação Evolucionária para a obtenção da sequência de execução de métodos para processamento de imagens e seus respectivos parâmetros com vistas à solução de um problema que envolva processamento de imagens. Este algoritmo estende alguns princípios de computação evolucionária ao introduzir a noção de níveis de otimização (local e global), com correspondentes estruturas e métodos para a representação da população e, mesmo direcionado para métodos para processamento de imagens, pode ser estendido para a solução de outros problemas. Para tanto, basta ter-se uma referência (*groundtruth*) e funções com as quais se acredita ser possível resolver ou aproximar uma solução para um problema específico.

O algoritmo desenvolvido foi capaz de produzir soluções aproximadas para problemas de alta complexidade computacional, e.g., a segmentação de imagens.

B.1 Definição da abordagem de computação evolutiva proposta

Para melhor compreensão do método proposto, foi construída, de maneira similar a outros algoritmos de Computação Evolucionária, uma metáfora com a biologia e os processos evolutivos.

Dado que um Algoritmo Genético (HOLLAND, 1992) realiza a evolução de uma população, pode-se afirmar que a população evoluída corresponde ao conceito biológico de uma espécie. Nesta pesquisa, além de evoluir uma dada espécie para encontrar o melhor indivíduo (ou conjunto de indivíduos) que pode resolver ou aproximar um dado problema, também se deseja saber qual a melhor espécie, ou seja, qual a sequência de funções que pode produzir ou aproximar a melhor solução. A aproximação realizada pelos AG pode ser obtida a partir da maximização ou da minimização de uma função objetivo. Neste apêndice, de modo a simplificar a explicação, serão apresentados apenas os cálculos destinados à maximização, sendo os cálculos destinados à minimização análogos aos apresentados.

Assim, sendo uma Espécie S definida por uma sequência F de n funções f_1, f_2, \dots, f_n e por um genoma g que corresponde ao conjunto de parâmetros necessários às n funções, a evolução consiste em encontrar a melhor Espécie e seu respectivo genoma para a solução do problema.

Portanto, seja um Gênero G um conjunto de m Espécies. Para a i -ésima espécie S_i , pertencente a G , busca-se o melhor genoma g_{melhor}^i da espécie dentre o conjunto de todos os genomas possíveis γ , que maximize (Equação B.1) a função objetivo f_{obj} , de acordo com a natureza do problema a ser solucionado.

$$g_{melhor}^i = \arg \max_g f_{obj}(g), g \in \gamma \quad (B.1)$$

Ao se comparar diferentes espécies, compara-se, de fato, a adequação do melhor genoma de uma espécie com o melhor genoma da segunda espécie. Assim, duas espécies

são equivalentes quando:

$$S_a \leftrightarrow S_b \Leftrightarrow f_{obj}(g_{melhor}^a) = f_{obj}(g_{melhor}^b) \quad (B.2)$$

Por sua vez, uma espécie é dita melhor do que a outra quando o melhor indivíduo (também denominado de genoma ou cromossomo) da primeira produz resultados melhores que o melhor indivíduo da segunda, conforme explicitado na Equação B.3.

$$S_a > S_b \Leftrightarrow f_{obj}(g_{melhor}^a) > f_{obj}(g_{melhor}^b), \quad (B.3)$$

sendo a espécie a melhor do que a espécie b , no caso de uma função de maximização ou b melhor do que a , no caso de uma função de minimização.

Assim, a melhor Espécie S_{melhor} é aquela que supera as demais Espécies existentes na solução específica do problema, conforme explicitado nas Equações B.4, para a maximização da função objetivo respectivamente.

$$S_{melhor} = s | \forall s, s' \in G \Rightarrow s \geq s' \quad (B.4)$$

Pode-se perceber que o processo de otimização do sistema consiste em dois níveis de otimização: no primeiro nível, mais interno, descobre-se qual o genoma que produz melhores resultados para a sequência de funções que define a espécie, enquanto no segundo nível, mais externo, descobre-se qual a espécie que, dentre aquelas existentes, produz os melhores resultados.

Define-se que duas Espécies são iguais se e somente se as sequências de funções que as definem forem as mesmas, como mostrado na Equação B.5.

$$F_a = F_b \Leftrightarrow |F_a| = |F_b| \wedge f_n^a = f_n^b \forall n \in |n| < |F_a| \quad (B.5)$$

$$S_a = S_b \Leftrightarrow F_a = F_b \quad (B.6)$$

em que F_a e F_b são as sequências de funções que definem as espécies S_a e S_b , respectivamente e f_n^a é a n -ésima função da sequência F_a .

Vale enfatizar que, por se tratar de uma sequência e não de um conjunto, duas

Espécies diferentes podem ser compostas pelas mesmas funções, embora dispostas em sequências distintas.

Uma Espécie possui uma população de n indivíduos, sendo $n > 0$. Os indivíduos são caracterizados por genomas. Dois indivíduos são diferentes quando possuem genomas diferentes. A igualdade das espécies é independente de seus genomas. Assim, mesmo com genomas diferentes, duas espécies podem ser iguais, desde que apresentem a mesma sequência de funções.

B.1.1 Surgimento das espécies

Dado que uma Espécie é definida por um genoma g e uma sequência de funções F , é preciso entender a formação destes dois entes.

A sequência F de funções que definem uma espécie é obtida partindo-se de um repositório de funções F_{rep} tais que $|F_{rep}| > 0$, com as quais se acredita ser possível solucionar um dado problema. Para que seja definida uma sequência válida de funções, é primordial que algumas regras sejam satisfeitas. Tais regras são explicitadas na Equação B.7:

$$\forall f \in F_{rep} \exists F' = \{f_1, f_2, \dots, f_n\} \subset F_{rep} | f_{ent} \circ f_1 \circ f_2 \circ \dots \circ f_n \circ f_{saida} \quad (B.7)$$

em que f_{ent} é uma função que gera dados de entrada a partir da função objetivo e f_{saida} a função que gera a avaliação numérica final para a função objetivo. A função *valida* retorna Verdadeiro se e somente se a composição que estiver sendo avaliada for possível.

Se a Equação B.7 for satisfeita, pode-se afirmar que, para qualquer função pertencente ao repositório, existe pelo menos uma sequência de funções que, ao se compor com a função f_{ent} , permitirá a obtenção de uma solução para a função f_{saida} .

Com relação à quantidade de sequências que podem ser geradas a partir de um repositório F_{rep} com $|F_{rep}| > 1$, considerando não ser possível que uma mesma função apareça em mais de uma posição na sequência F , a quantidade de combinações $Comb$

possíveis é:

$$Comb = \sum_{i=1}^{|F_{rep}|} C_{|F_{rep}|}^i \quad (B.8)$$

Caso sejam admitidas repetições de funções na sequência e, em se admitindo repetições de funções na sequência, e não havendo um número máximo de combinações, existirão infinitas combinações de tais funções.

Para evitar que a quantidade de possibilidades resulte em uma explosão de combinações, são definidas algumas restrições de criação de espécies:

1. Não pode existir mais de uma espécie com a mesma configuração de sequência de funções, ou seja, todas as espécies são diferentes;
2. As sequências de funções que definem as espécies serão criadas com o menor tamanho possível e de modo a gerar uma saída válida para o problema a partir da entrada fornecida pela função objetivo.

Até o presente momento, não foi desenvolvido um método inteligente para a construção da sequência de funções, a qual tem sido realizada por meio de funções aleatórias. Dentre as investigações a serem desenvolvidas estão a de utilizar estruturas de dados mais complexas, como, por exemplo, Árvores, para que seja possível identificar a raiz de operações que possui as combinações mais promissoras, de modo a balancear a criação de novas espécies, a partir de informações anteriormente adquiridas.

B.1.2 Evolução e sobrevivência das espécies

A quantidade inicial de espécies que serão criadas é definida no momento da inicialização do Gênero. Uma vez criadas as m espécies, inicia-se a evolução. Cada Espécie evolui isoladamente de maneira a encontrar, para a sequência de funções que a define, o melhor genoma. Todas as espécies evoluem a mesma quantidade de gerações. Todas as espécies são inicializadas com a população de mesmo tamanho. Para a evolução de uma Espécie, são utilizados Algoritmos Genéticos.

Ao final da evolução realizada pelos Algoritmos Genéticos de cada Espécie, tem-se o desempenho geral da Espécie para o mesmo problema e o mesmo número de gerações. Desta forma, pode-se indicar qual das Espécies teve melhor desempenho

para o problema indicado. Entretanto, este resultado apresenta apenas o desempenho geral das m espécies criadas. Outras espécies ainda não criadas podem ter desempenho melhor do que as espécies atualmente criadas. Portanto, para permitir que outras espécies sejam criadas e comparadas às Espécies atuais, deverá ser iniciado um novo ciclo, denominado, nesta pesquisa, de Era.

No início de cada Era, x novas Espécies são criadas. As Espécies existentes continuam o treinamento interrompido ao final da Era anterior. Esse comportamento deve ocorrer para que as novas Espécies sejam comparadas àquelas ora existentes, a fim de que se possa avaliar a adequação das diferentes Espécies ao problema. Este processo é repetido até que um número máximo de eras é atingido.

Naturalmente, este processo implica um alto custo computacional, pois exige o processamento de diversas evoluções que utilizam Algoritmos Genéticos, os quais são processos computacionalmente onerosos. Para minimizar este problema, foi definido que a população de cada Espécie sofrerá variação a cada Era. Assim, as Espécies que não obtiverem êxito em suas funções de adaptação terão suas populações progressivamente reduzidas até serem extintas, caso continuem apresentando mau desempenho. Por outro lado, as Espécies que conseguirem melhorar seus desempenhos terão suas populações aumentadas. As funções para o cálculo populacional das espécies são apresentadas nas Equações B.9 a B.11.

$$f_{natalidade}(k) = \frac{f_{obj}(g_{melhor}^{S_i}, k)}{f_{obj}(g_{melhor}^{S_i}, k-1)}, \quad (B.9)$$

$$f_{mortalidade}(k) = \frac{f_{obj}(g_{melhor}^{S_{melhor}}, k)}{f_{obj}(g_{melhor}^{S_i}, k)}, \quad (B.10)$$

$$f_{crescimento}(k) = f_{natalidade}(k) - f_{mortalidade}(k), \quad (B.11)$$

em que $g_{melhor}^{S_i}$ representa o melhor genoma da i -ésima Espécie, $g_{melhor}^{S_{melhor}}$ representa o melhor genoma da melhor Espécie encontrada e k representa a Era atual. Vale ressaltar que o melhor genoma de uma espécie é o melhor genoma encontrado em todo o tempo em que a espécie manteve-se viva e não apenas na última Era, o que implica $f_{natalidade} \geq 1$ e, similarmente, $f_{mortalidade} \geq 1$. Portanto, o cálculo da população P da

Espécie i para a Era $k + 1$ pode ser realizado a partir da Equação B.12.

$$P_{i,k+1} = P_{i,k} \times (1 + f_{\text{crescimento}}) \quad (\text{B.12})$$

O método de crescimento populacional é, provavelmente, a chave para que um bom treinamento seja realizado sem que haja explosão combinatorial. Uma proposta de melhoria envolve a utilização de outros tipos de funções para o cálculo da população.

Ainda não foi desenvolvido um método automático para o encerramento do processo de treinamento. Até o presente estágio, o treinamento encerra-se no momento em que a quantidade de Eras atinge o máximo permitido. Como proposta de melhoria, podem-se utilizar os resultados obtidos para identificar se devem ser criadas novas sequências a partir de funções que não apresentaram bons resultados.

Para os testes do método proposto, foi efetuada uma variação na biblioteca GAlib (WALL, 2009). O processo de adaptação é descrito na Seção B.2 deste apêndice.

B.1.3 Comparação com os algoritmos genéticos

Duas questões surgem naturalmente quanto ao método proposto: (1) Não se trata de um método equivalente aos Algoritmos Genéticos tradicionais? e (2) Não é possível, com o mesmo desempenho e facilidade de programação associados aos Algoritmos Genéticos tradicionais, produzir o mesmo resultado?

Para facilitar a discussão, é necessário considerar o seguinte problema hipotético. Sejam f_1, \dots, f_n funções das quais se dispõem. Deseja-se testar: (1) qual a combinação de tais funções produzem melhores resultados; e (2) quais os parâmetros dessa combinação que também resultam em melhores resultados. Vale ressaltar que algumas das funções podem não ser utilizadas, enquanto outras podem ser utilizadas mais de uma vez.

Para o problema proposto, considere-se que cada função possui uma quantidade de parâmetros $n_p \geq 1$. Suponha-se uma dada distribuição de cinco funções, f_1, f_2, f_3, f_4 e f_5 e a respectiva quantidade de parâmetros que cada função necessita, conforme apresentado na Tabela B.1.

Tabela B.1: Funções e respectivas quantidades de parâmetros.

f_1	3
f_2	1
f_3	4
f_4	1
f_5	5

Para representar tais funções, pode ser criado um genoma, para o qual o primeiro fenótipo, do tipo inteiro, representa o índice da função e os demais fenótipos, do tipo Real, representam os respectivos parâmetros de cada função. Para tanto, cada genoma pode ter dimensões bem diferentes. Logo, uma primeira consideração para a realização desta evolução é:

Consideração 1: Existem abordagens de Algoritmos Genéticos que não suportam evoluções de genomas de tamanhos diferentes.

Na abordagem Canônica de Algoritmos Genéticos, proposta por Holland (HOLLAND, 1992), por exemplo, não está previsto o uso de genomas de tamanhos diferentes. O uso de genomas de tamanhos diferentes é possível em detrimento de cruzamentos mais complexos.

Como exemplo de dificuldades causadas por genomas de tamanhos diferentes, considere-se que a abordagem utilizada suporte esse tipo de evolução. No início da evolução, podem-se formar três diferentes indivíduos (equivalentes a três espécies no algoritmo proposto), os quais são representados da seguinte forma:

$$S_a = f_1(\theta_1, \theta_2, \theta_3)$$

$$S_b = f_2(\theta_1)f_1(\theta_2, \theta_3, \theta_4)$$

$$S_c = f_1(\theta_1, \theta_2, \theta_3)f_4(\theta_4)f_1(\theta_5, \theta_6, \theta_7),$$

sendo S_a uma espécie composta apenas pela função f_1 e seus três parâmetros (θ_1 , θ_2 e θ_3), S_b uma espécie composta pelas funções f_2 e f_1 , nesta ordem, contando, ao todo, com 4 parâmetros (1 parâmetro da função f_2 e 3 parâmetros da função f_1) e, analogamente, S_c composta pelas funções f_1 , f_4 e f_1 , nesta ordem, contando, ao todo, com 7 parâmetros (θ_1 a θ_n).

No momento da execução da função objetivo, uma segunda consideração pode ser levantada:

Consideração 2: Como realizar cruzamentos entre indivíduos de formatos diferentes?

Conforme mencionado, o problema do cruzamento precisa ser resolvido para casos em que os genomas possuem tamanhos diferentes. Assim, para uma melhor visualização do problema, será ilustrado o cruzamento entre S_a e S_b , utilizando-se o *One-Point Crossover*, método utilizado na combinação entre os indivíduos (observe que as numerações dos parâmetros foi modificada, para melhor ilustração do problema):

$$Pai = S_a = f_1(\theta_1, \theta_2, \theta_3)$$

$$Mae = S_b = f_2(\theta_4)f_1(\theta_5, \theta_6, \theta_7)$$

$$Filho = f_1(\theta_1, \theta_2, \theta_6)\theta_7$$

$$Filha = f_2(\theta_4)f_1(\theta_5, \theta_3,)$$

Em um cruzamento desta natureza, ambos os filhos gerados não apresentam uma representação completa das funções. No caso do *Filho*, sobra um parâmetro; no caso da *Filha*, falta um parâmetro.

Contudo, este problema pode ser amenizado, caso o método de cruzamento seja capaz de prever situações desta natureza e garantir que a permutação de valores respeite os números de parâmetros. Pode-se, assim, produzir os seguintes filhos válidos:

$$Filho = f_1(\theta_4, \theta_2, \theta_3)$$

$$Filha = f_2(\theta_1)f_1(\theta_5, \theta_6, \theta_7)$$

Na prática, os significados, tipo e faixa de valores dos parâmetros θ_1 e θ_4 podem ser diferentes. Para evitar inconsistências, poder-se-ia usar números reais compreendidos na faixa [0..1]. A consequência é o drástico aumento na faixa de busca. Neste caso, a informação que pode ser representada em um bit, por exemplo, pode ter a representação de um número Real, o que pode requerer, por exemplo, 32 bits, dependendo da representação considerada.

Outra forma de solucionar este problema é, no momento do cruzamento, apenas recombinar parâmetros de funções equivalentes, o que impedirá o cruzamento intra-espécies, requerendo o conhecimento prévio das estruturas dos indivíduos, o que substituiu uma das características da abordagem proposta.

Consideração 3: Como verificar a compatibilidade entre funções?

Assumindo que o cruzamento pode ser solucionado a partir de alguma função de

cruzamento e que todos os parâmetros sejam equivalentes, surge um novo problema: como verificar se a sequência de funções resultante do cruzamento representa uma função válida?

Para que um resultado seja encontrado, é preciso que a saída das funções intermediárias sejam compatíveis com as entradas de suas subsequentes. Por exemplo, uma função pode ter como resposta valores do tipo Real, enquanto sua subsequente requer valores do tipo Booleano.

Uma vez que se sabe *a priori* a compatibilidade entre as funções existentes e se sabe a sequência das funções associada ao cruzamento, pode-se avaliar, para cada indivíduo gerado, se este é formado por uma sequência válida de funções. Apesar de ser uma abordagem teoricamente possível, pode ser impraticável, uma vez que tal verificação terá que ser executada a cada execução da função objetivo ou, opcionalmente, da função de cruzamento. Dependendo da diversidade da população, pode inclusive, não ser possível a geração de nenhum indivíduo válido.

De certa forma, todas as considerações elencadas podem ser resolvidas de alguma das formas supracitadas ou de outras formas. Entretanto, pode haver desperdício de recursos computacionais nos processos de validação e adaptação das informações.

Adicionalmente, a informação da sequência de funções e respectivos parâmetros está agrupada em uma mesma sequência de informações, o genoma. Uma vez que a abordagem proposta desacopla a combinação das funções de seus parâmetros, é possível realizar um treinamento mais efetivo, à medida que (1) a sequência das funções e os parâmetros necessários para a execução daquela combinação de funções serão re-combinados isoladamente; (2) será utilizada a quantidade exata de bits para a representação da informação; e (3) não há necessidade de validação da sequência a cada indivíduo gerado, uma vez que a validação ocorre apenas na criação da espécie.

B.1.4 Programação genética

Uma extensão dos algoritmos genéticos que também pode ser comparada à presente pesquisa é a programação genética (GP) (KOZA, 1992). A GP tem por objetivo evoluir programas. Portanto, comparativamente aos GA, os indivíduos são programas e estes se recombinaem dando origem a novos programas. Os cromossomos são sequências de

operações lógico-aritméticas as quais comumente são organizadas em formato de árvore.

Assim como na comparação com os GA, não é possível diferenciar funcionalmente a presente proposta da GP. Entretanto, podem-se identificar alguns benefícios da utilização da presente proposta. Inicialmente, pode-se identificar a estruturação realizada sobre espécies e gêneros, que simplifica a compreensão da extensão dos GA. Adicionalmente, a presente investigação foi estruturada em função de algoritmos de processamento de imagens e procedimentos mais complexos que recebem diversos parâmetros, o que requereu um tratamento quanto à compatibilidade entre as entradas e saídas das funções (e.g., uma função que processa uma imagem no espaço de cor HSV não pode receber uma imagem que esteja no espaço de cor RGB).

B.1.5 Algoritmo de segmentação

Para identificar a sequência de processamento utilizada no processo de segmentação das imagens, foram utilizados os seguintes algoritmos:

- **Filtro suavizador.** Como a implementação utilizada é a disponível na biblioteca OpenCV (BRADSKI, 2000), foram testadas as seguintes variações: suavização, suavização sem escala, filtro da mediana, filtro gaussiano e filtro bilateral. A suavização pode ser útil no processo de segmentação uma vez que possibilita a homogeneização das regiões de cor;
- **Filtro homomórfico.** O filtro homomórfico pode ser útil no processo de segmentação, uma vez que pode reduzir efeitos de iluminação. Foram incorporadas duas implementações do filtro homomórfico, uma para o espaço de cor HSV e outra para o espaço de cor LAB, cuja única diferença é o canal de intensidades a ser homogeneizado;
- **Mean-shift.** O algoritmo de segmentação *mean-shift* (COMANICIU; MEER, 2002) pode ser utilizado para segmentar regiões na imagem;
- **Agrupador HSV.** Este algoritmo, descrito na Seção B.1.6, pode ser utilizado para demarcar as principais regiões da imagem e suas respectivas cores dominantes;

- **Conversão bidirecional entre espaços de cor RGB, HSV e LAB.** Algoritmo destinado à conversão simples de um espaço de cor para o outro, empregando a implementação do OpenCV (BRADSKI, 2000);

A partir destes algoritmos, foi desenvolvida uma estratégia para produzir sequências de processamento, a qual consistiu na pré-seleção de funções compatíveis com cada tipo de dado de entrada considerado. Posteriormente, uma das funções compatíveis é escolhida aleatoriamente. Este procedimento é repetido até que a última função incorporada à sequência seja compatível com o tipo de dado de saída e desde que a sequência de funções não tenha sido previamente produzida, conforme explicitado na Equação B.7.

Um dos algoritmos passíveis de combinação, o agrupador HSV, será descrito na Seção B.1.6.

B.1.6 Estratégia para agrupamento a partir do espaço de cor HSV

A estratégia para o agrupamento com base no espaço de cor HSV é inspirado no algoritmo *k-means*, sem que, contudo, seja necessária a indicação da quantidade de núcleos. Tal característica é importante, uma vez que não se sabe *a priori* a quantidade de núcleos de cor que a imagem terá.

A estratégia consiste em (1) analisar o histograma de matizes da imagem para a identificação de possíveis núcleos de cor; (2) identificar, para cada *pixel* da imagem, qual o núcleo de cor do qual o pixel considerado mais se aproxima; e (3) verificar, neste novo conjunto de *pixels*, se existe alguma cor que esteja mais próxima às demais cores do que o núcleo de cor inicialmente indicado. Para acelerar o processamento, apenas uma iteração é feita.

B.2 Aspectos computacionais

A implementação do método apresentado na Seção B.1 levou em consideração dois principais fatores. O primeiro foi a eficiência computacional, uma vez que soluções

envolvendo AG costumam ser lentas, a execução de vários AG poderia inviabilizar o treinamento. O segundo fator considerado é a flexibilidade. Ainda que, neste capítulo, o método proposto tenha sido utilizado para a solução de um problema de processamento de imagens, tal método pode permitir que problemas de outras naturezas sejam solucionados. Para que um problema possa ser resolvido utilizando o método proposto, é necessário que o problema (1) seja solucionável por intermédio de composição de funções, (2) seja supervisionado, logo é preciso ter um conjunto de entradas e as saídas desejadas e (3) tenha natureza de otimização.

Sendo assim, para amenizar o primeiro fator, o sistema foi desenvolvido utilizando-se *threads* em C++. A consequência é que, em processadores multi-núcleo, tantas espécies quantos forem os núcleos podem ser processadas simultaneamente. Além disso, ficará mais simples estender o sistema para um *grid* ou um *cluster*, sendo cada nó responsável pelo processamento de um número de espécies que independe do restante do processamento. A sincronização, neste caso, é feita ao final de cada Era.

B.3 Adaptação da biblioteca GALib

Para que fosse possível a utilização de *threads*, foi necessária uma adaptação na biblioteca GALib (WALL, 2009). A referida biblioteca possui um módulo gerador de números aleatórios (GNA) que utiliza variáveis globais ao sistema. Para simular a aleatoriedade, após a leitura de um número aleatório, tais variáveis globais são modificadas. Uma vez que durante a geração dos números aleatórios, faz-se busca a algumas posições de memória ocupadas por estas variáveis, na execução paralela esse comportamento frequentemente leva à interrupção abrupta da execução, que ocorre por acesso inválido quando, anteriormente ao acesso de uma posição de memória por uma *thread* em execução, a variável global é alterada por outra *thread* em execução.

Uma das possíveis soluções, seria a utilização de semáforos, protegendo a variável de escrita enquanto esta estivesse sendo lida. Entretanto, como o módulo de GNA é utilizado em diversos setores do código, a utilização de semáforos praticamente eliminaria a possibilidade de paralelismo, pois uma leitura é quase sempre acompanhada de uma escrita. Assim, para resolver estes problemas, foi construída uma classe destinada

à Geração de Números Aleatórios. Assim, cada espécie tem sua própria classe geradora, com suas próprias variáveis. A adaptação da biblioteca GALib (WALL, 2009) foi realizada, de modo a substituir todas as chamadas estáticas e globais para chamadas de métodos de objetos que são instanciados apenas no momento da inicialização.

O diagrama de classe da abordagem de computação evolutiva descrita é apresentado na Figura B.1.

Portanto, para o desenvolvimento de um sistema que efetue o treinamento mencionado neste capítulo, é preciso realizar as interfaces *GeradorDeEspécies* e *CompetidorDeEspécies* e implementar a classe abstrata *Espécie* a partir do desenvolvimento do método *funçãoObjetivo*, responsável pela avaliação de um indivíduo da espécie.

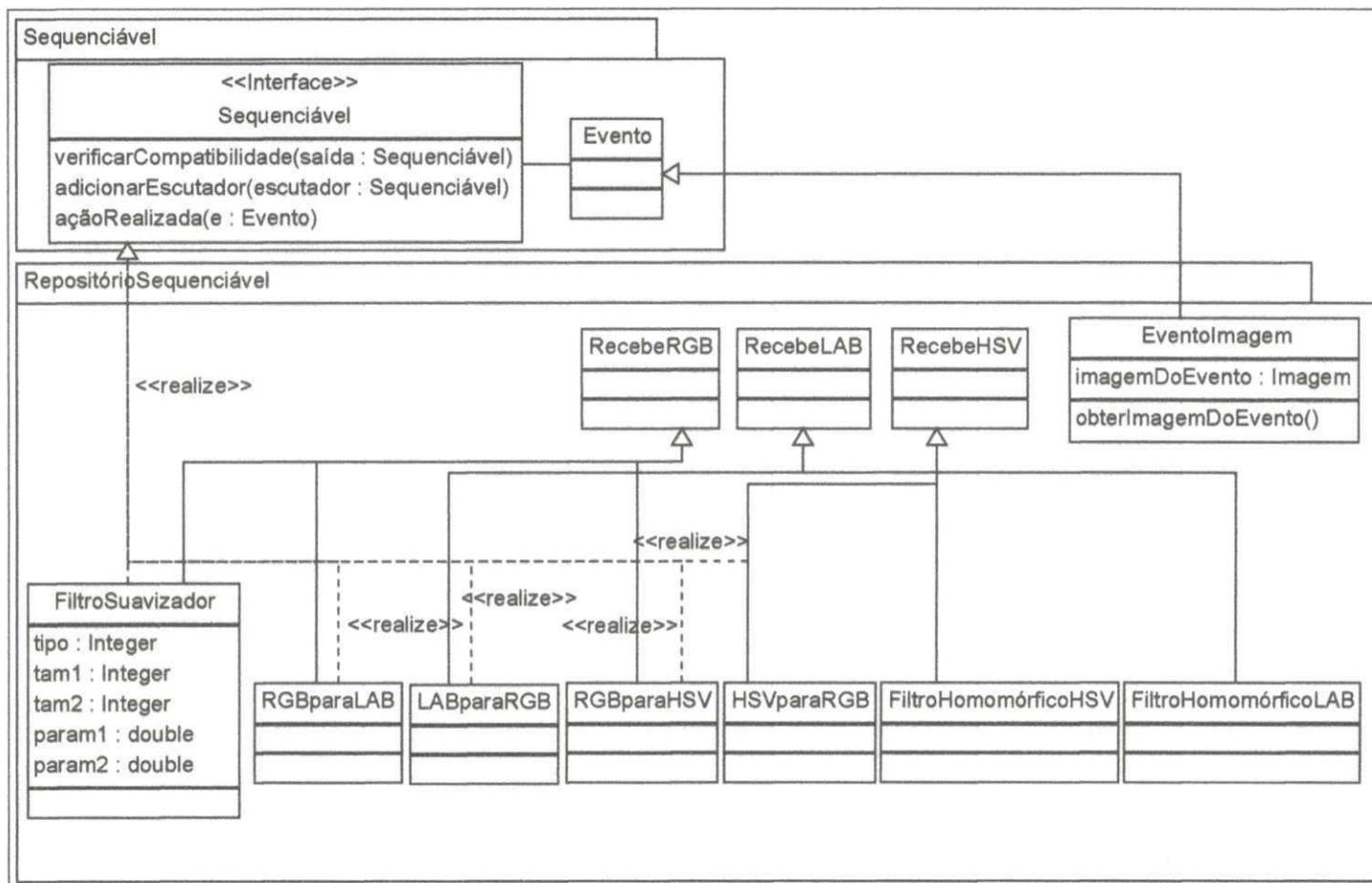
O controle da evolução das espécies se dá pela classe *Gênero* por intermédio do método *evoluir*, o qual é o comando para início do treinamento e que permanece em execução até o término do treinamento.

O modelo para o processamento sequencial das estruturas de classe apresentada na Figura B.1, é apresentado na Figura B.2. Nesta tese, o processamento sequencial serve para avaliar se um determinado módulo de processamento poderá se conectar em série com um outro módulo de processamento.

Portanto, para que uma classe faça parte do conjunto de classes destinadas ao processamento sequencial, é preciso que sejam implementados os métodos abstratos *verificarCompatibilidade*, *adicionarEscutador* e *açãoRealizada*. Na Linguagem de Programação C++, a solução para implementação desta funcionalidade é por intermédio do procedimento *dynamic_cast*. O procedimento *dynamic_cast* realiza conversão de tipo em tempo de execução. Para que a conversão seja realizada com êxito a partir do *dynamic_cast* é preciso que o objeto seja derivado da classe base do ponteiro ao qual se deseja converter, caso contrário a conversão resultará em um ponteiro nulo. Assim, por exemplo, uma classe que seja derivada da classe *RGBParaLAB* poderá ser convertida, por intermédio do *dynamic_cast*, para um ponteiro do tipo *RGBParaLAB*. Por intermédio desta abordagem, é possível saber se uma determinada classe possui os métodos relativos a um determinado comportamento (e.g., ser capaz de tratar imagens no espaço de cor HSV), permitindo a conectividade entre classes em tempo de execução.

A Figura B.2 também contém o modelo de algumas classes utilizadas nesta pesquisa para a segmentação de imagens, e.g., Filtro Suavizador, Conversor do Espaço de cor RGB para os espaços de cor LAB e HSV e Filtro Homomórfico. As classes utilizadas no processo de segmentação são descritas em mais detalhes na Seção B.1.5.

Figura B.2: Modelo de processamento sequencial utilizado.



B.4 Considerações finais

Durante a produção desta tese, em particular durante a produção de um método para a segmentação automática de fotografias, foi necessário o desenvolvimento de um método para testar, de forma metodológica, uma grande quantidade de possíveis soluções.

Uma vez que as ferramentas atuais não davam suporte simples a esta tarefa, foi desenvolvido um arcabouço que, utilizando AG, combina as ferramentas de processamento de imagens desenvolvidas, facilitando e automatizando do processo de obtenção da solução.

Apêndice C

Evolução da tecnologia de captura de fotografias

Parte da motivação desta pesquisa advém da necessidade crescente de se processar grande volume de fotografias digitais. Uma vez que estudos publicados sobre a evolução dos mercados de fotografia digital são, muitas vezes, inacessíveis, torna-se difícil avaliar qual o impacto desta pesquisa em tal mercado.

Assim, compilou-se neste apêndice um breve estudo sobre a evolução da fotografia digital nos últimos anos, partindo de 1975, quando do advento da primeira câmera digital, passando pelos dias atuais (lembrando que este documento foi redigido no período compreendido entre 2009 e 2011) e vislumbrado as perspectivas nesta área para os próximos anos.

Não se objetiva neste apêndice um estudo completo do assunto, mas tão somente a apresentação de um breve levantamento, com base em reportagens recentemente publicadas sobre o tema. A pesquisa também se limita ao estudo de câmeras digitais destinadas a consumidores comuns (ou amadores), não sendo voltada para câmeras profissionais ou para dispositivos específicos, tais como sensores a bordo de plataformas de satélites ou câmeras para microscopia. Busca-se apresentar uma breve introdução sobre perspectivas e tendências que poderão contribuir para o melhor entendimento da área de pesquisa.

Este apêndice foi estruturado como segue. Na Seção C.1, relata-se sucintamente a evolução tecnológica pela qual passaram as câmeras digitais. Na Seção C.2, discutem-

se detalhes relativos à variação dos preços das câmeras digitais ao longo dos últimos 10 anos. Por sua vez, na Seção C.3, são discutidas as mudanças ocorridas no processo de obtenção e revelação de fotografias sob a ótica do consumidor. Por fim, na Seção C.4, são formuladas algumas considerações finais com base no exposto neste Apêndice.

C.1 Evolução da tecnologia das câmeras digitais

Existe o registro do primeiro protótipo de uma Câmera Digital ter sido desenvolvido em 1975, pelo projetista da Kodak Steven Sasson. A câmera em questão capturava imagens monocromáticas de dimensões 100x100 *pixels* e pesava quase 4 quilogramas, requerendo 23 segundos para a gravação de uma única fotografia em uma fita cassete (GREENSPUN, 2007; AYRES, 2007; CARTER, 2001).

Apesar do processo fotográfico digital ter se alterado relativamente pouco de 1975 até os dias atuais, o projeto da câmera fotográfica digital sofreu alterações dramáticas no referido período. As dimensões das câmeras, por exemplo, foram, via de regra e grosso modo, reduzidas. É pertinente ressaltar que, em alguns modelos, o tipo de lente empregado ainda pode requerer grandes dimensões do corpo da câmera. Entretanto, surgiram diversas câmeras compactas, passíveis de acoplamento a dispositivos tais como celulares e *notebooks*.

A segunda mudança ocorreu nos mecanismos de correção automática das configurações da câmera, tais como foco automático, correção do balanço do branco, ajuste do tempo de exposição e abertura do diafragma, dentre outros. Tal mudança teve um impacto significativo na popularização da fotografia, uma vez que passou a permitir que indivíduos sem nenhuma experiência ou conhecimento dos conceitos básicos de fotografia pudessem fotografar.

Outro fator de destaque reside na tecnologia e capacidade de armazenamento das câmeras fotográficas digitais. Enquanto no modelo proposto por Sasson eram utilizadas fitas cassete, o modelo Sony Mavica utilizava discos flexíveis de 1,44MB (CARTER, 2001). Posteriormente, muitas câmeras passaram a conter dispositivos de armazenamento em estado sólido internos, os quais requeriam cabos para a transferência das fotografias para o computador. Atualmente, a tecnologia mais utilizada é a de memó-

rias eletrônicas (e.g., SD, SDxC, Compact Flash, dentre uma série de modelos).

Deve-se também destacar a evolução na resolução da captura. De 0,01 *Megapixels* da câmera proposta por Sasson até câmeras com mais de 24 *Megapixels* nos dias atuais. Parte dos modelos de elevada resolução ainda é destinado a uma clientela específica de profissionais de fotografia, fora do escopo desta pesquisa.

Por fim, uma das possibilidades trazidas pela tecnologia digital foi a capacidade de visualização da fotografia sem a necessidade do processo de revelação. A presença de um visor colorido permitiu uma resposta rápida e de boa qualidade, tornando possível identificar pequenas imperfeições na fotografia e, quando necessário, capturar uma nova fotografia em questão de frações de segundo. Além da visualização na própria câmera, a possibilidade de se transferir a fotografia diretamente para um computador, sem a necessidade de aparelhos digitalizadores, trouxe novos usos para a fotografia digital. tais como o envio por correio eletrônico, a impressão doméstica e, mais recentemente, a publicação em redes sociais.

C.2 Evolução dos preços das câmeras digitais

Dado que o ato de fotografar requer equipamentos, faz-se necessário estudar como se comportou o custo para a sua obtenção neste período de tempo.

Naturalmente, os preços da maioria dos produtos sofrem influências de leis de mercado. Sendo assim, novos modelos e inovações tecnológicas, durante os primeiros meses após o lançamento, atingem preços mais elevados do que os produtos ora existentes e, aos poucos, vão sofrendo reduções de preço. Os preços são também influenciados pela concorrência e produção em massa por parte dos fabricantes (DUNLEAVY, 2006).

Não constitui uma tarefa trivial analisar a evolução dos preços das câmeras digitais. O principal motivo é que são vários os fatores que compõem o preço de uma câmera: marca, quantidade de *Megapixels*, número de fotografias por segundo, opções extra tais como capacidade de gravar vídeos, *design*, peso, capacidade da bateria, potência do *flash*, tamanho do visor digital, tipo e tamanho das lentes e processador, dentre outros.

A Tabela C.1 contém alguns modelos de câmeras, sendo destacados o ano de lançamento, a quantidade de *Megapixels* e a quantidade de fotografias por segundo que a

câmera é capaz de obter. Os preços estão em Dólares Americanos, MP é a quantidade de *Megapixels* e FPS é a quantidade de *quadros* por segundo no modo rajada (*burst*).

Tabela C.1: Tabela comparativa de preços de câmeras digitais da marca Nikon, entre 1999 e 2010, conforme registrados por Rockewll (2011).

Ano	Modelo	Preço	MP	FPS	Ano	Modelo	Preço	MP	FPS
1999	D1	\$5000	2.7	4.5	2007	D300	\$1800	12	8
2001	D1H	\$5000	2.7	5	2007	D3	\$5000	12.1	9
2001	D1X	\$5000	5.3	3	2008	D700	\$2999	12	5
2002	D100	\$1999	6	3	2008	D90	\$999	12	4.5
2003	D2H	\$5000	4.1	8	2008	D3X	\$7500	24.4	5
2004	D70	\$999	6	3	2009	D5000	\$730	12	4
2004	D2X	\$5000	12	5	2009	D3000	\$599	10	3
2005	D2Hs	\$3500	4.1	8	2009	D300s	\$1799	12	7
2005	D50	\$1299	6	2.5	2009	D3s	\$5200	12.1	9
2005	D200	\$1699	10	5	2010	D3100	\$699	14	3
2006	D2Xs	\$4700	12	5	2010	D7000	\$1195	16	6
2006	D80	\$999	10	3	2010	D5100	\$799	16	4

Para tentar realizar esta comparação, foram feitas duas restrições. A primeira restrição diz respeito à marca comparada, que será sempre a mesma (Nikon). Rockewll (2011) realizou um registro dos principais e mais populares modelos digitais ora lançados pela marca Nikon, sendo tal registro utilizado nesta pesquisa para a avaliação da evolução dos preços. A segunda restrição foi quanto à data do registro. Em algumas situações, Rockewll (2011) cita preços do mesmo modelo de câmera em épocas diferentes. Assim, para evitar a duplicação de registros de preço, sempre foi utilizado o registro de preço mais antigo.

Conforme exibido na Tabela C.1, os preços podem variar drasticamente de acordo com o modelo da câmera. Entretanto, dentro de uma mesma família de modelos, a variação é bem menor, e.g., os modelos D1, D1H, D1X, D2, D2s, D3 e D3s. Portanto, para melhor observar a evolução tecnológica, é mais pertinente a avaliação de uma mesma família de modelos.

Para efetuar tal comparação, foram consideradas as câmeras da família D100 D300s (ver Tabela C.2). Os preços estão em Dólares Americanos, MP é a quantidade de *Megapixels* e FPS é a quantidade de quadros por segundo e PPMP é o preço por *Megapixel*.

Tabela C.2: Tabela comparativa dos preços de câmeras digitais da família D100 a D300s, conforme registrados por Rockewll (2011).

Ano	Modelo	Preço	MP	FPS	PPMP
2002	D100	\$1999	6	3	\$333,16
2005	D200	\$1699	10	5	\$169,90
2007	D300	\$1800	12	8	\$150,00
2009	D300s	\$1799	12	7	\$149,92

Pode-se perceber, ao analisar a Tabela C.2, que o preço por *Megapixel* sofreu decréscimo de aproximadamente 55% entre 2002 e 2009. Vale ressaltar que esta variação do preço também está vinculada a outros componentes e características da câmera, e.g., o número de quadros por segundo.

Com relação aos demais equipamentos fotográficos, a oscilação do preço ocorreu de acordo com a receita dos grandes fabricantes. Pôde-se verificar que houve prejuízo para estes fabricantes no período compreendido entre 2005 e 2007. Uma vez que parte de tal prejuízo foi atribuída à concorrência de pequenos fabricantes de câmeras e à incorporação de câmeras aos telefones celulares, foi feito um direcionamento de mercado de modo a provocar um aumento no interesse em câmeras mais sofisticadas como as SLR digitais (SULLIVAN; MONSON, 2005).

Um efeito semelhante àquele supramencionado pode ser percebido atualmente. Com a melhoria das tecnologias de captura para dispositivos móveis, tais como celulares e *tablets*, criou-se o receio de que isso reduziria o número de fotografias obtidas a partir de câmeras fotográficas.

De fato, em comunidades virtuais como a Flickr.com, percebe-se que alguns dispositivos, especificamente o iPhone, estão chegando ao topo como dispositivos mais utilizados para envio de fotografias, o que não implica uma queda na quantidade de fotografias obtidas por outros dispositivos (CHEN, 2011).

Portanto, pode-se inferir com isso que os usuários costumam adotar dispositivos móveis ou câmeras mais simples para as fotografias do dia-a-dia (transformando a obtenção de fotografias em uma rotina quase diária), mas continuam utilizando câmeras mais sofisticadas para situações não rotineiras, tais como viagens e eventos especiais.

C.3 Mudanças no hábito de fotografar e no processo pós-fotografia

Uma das principais características que promoveram a popularização das câmeras fotográficas digitais foi a possibilidade de visualização e compartilhamento da fotografia sem que houvesse a necessidade de revelação da mesma.

Uma das principais consequências é a possibilidade de se obter mais fotografias. Em primeiro lugar, devido ao fato dos cartões de memória permitirem uma quantidade de fotografias significativamente superior àquelas que um rolo de filme comumente oferecia (i.e., 36). Em segundo lugar, porque não existe um custo atrelado à visualização da fotografia, e.g., o custo da revelação. Por fim, com a rápida resposta visual provida pela tela digital, pode-se decidir de imediato sobre a necessidade de descarte da fotografia feita e a aquisição de uma nova fotografia.

Portanto, por anos acreditou-se que as câmeras fotográficas digitais trariam prejuízo à indústria de impressão fotográfica, pois provocariam a redução do número de revelações. Apesar de números oficiais sobre tal fato não serem amplamente divulgados, pode-se perceber que na prática não foi o que aconteceu.

Em geral, a maior parte dos aficionados por fotografias opta pela montagem de álbuns sobre o evento no qual as fotografias foram obtidas, o que continuou sendo parte do processo fotográfico, com a diferença de que agora se poderia dispor de número ainda maior de fotografias para serem impressas e inseridas nos álbuns.

Para muitos, a necessidade da revelação era um fator que as afastava do hábito da fotografia. Para tais participantes, houve a possibilidade de se obter fotografias sem nenhuma obrigação adicional, tal como ter que (1) comprar um rolo de filme, (2) armazenar cuidadosamente o filme e levá-lo para revelação ou (3) esperar em média 1 hora para a obtenção das fotos, após o processo de revelação. Também deve ser levada em consideração a decepção associada ao processo: em muitos casos as fotografias “queimadas” (quando o tempo de exposição era maior do que o necessário ou houve exposição direta à luz), “escuras” (quando o tempo de exposição era menor do que o necessário) e “cortadas” (devido às câmeras mais simples possuírem um visor que não refletia a imagem a ser capturada pelo obturador, causando um efeito de paralaxe).

Por fim, as empresas de impressão de fotografias necessitavam de equipamentos caros e de difícil manutenção para a revelação dos filmes, além de mão-de-obra especializada para lidar com o processo fotográfico. A fotografia digital eliminou parte destes requerimentos, estimulando, inclusive, a construção de quiosques de auto-serviço para a revelação de fotografias digitais, o que passou a exigir somente a existência de um funcionário para substituição do material e eventuais manutenções. Vale ressaltar que o processo de transporte e armazenamento das fotografias também sofreu um impacto significativo, dado que as fotografias digitais podem ser transportadas em mídias tais como CD, DVD e Memórias *Flash*. Isso permite que o cliente possa pré-selecionar as fotografias que deseja imprimir, reduzindo o tempo que parte das empresas de revelação gastava para remover fotografias com problemas.

Diante do exposto, acredita-se que houve um aumento na procura pela revelação de fotografias e não uma queda, como inicialmente se esperava.

C.4 Conclusão

Constata-se que a fotografia digital é uma tecnologia que se beneficiou de várias evoluções tecnológicas que aconteceram na década de 90.

Percebe-se também que houve um aumento significativo nas vendas de câmeras SLR, em detrimento da venda de câmeras “*point-and-shot*”. Tal fato evidencia o crescimento do interesse por fotografias, a ponto de muitos chegarem a investir em equipamentos mais sofisticados e, por conseguinte, mais caros, os quais usualmente requerem conhecimentos mais aprofundados do ato de fotografar.

Esse aumento também pode ser percebido a partir das projeções favoráveis de aumento de receita, formuladas pela empresa de consultoria mercadológica In-stat, de acordo com sítio especializado CLN (2011), as quais preveem um crescimento nas receitas para o patamar de \$43.5 bilhões em 2015, estando 40% associados à venda de câmeras SLR.

Pode-se, portanto, concluir que o investimento em ferramentas para gerenciar a quantidade de fotografias que virão a surgir com esse aumento no maquinário é pertinente.

Apêndice D

Questionário sobre os hábitos fotográficos

Ao acessar a página do experimento descrito na Seção 7.2.2, solicitava-se ao participante que respondesse a um questionário sobre hábitos fotográficos. O objetivo deste questionário é, ao mesmo tempo em que se obtém a rotulagem sobre cores importantes, identificar o perfil de possíveis clientes para o sistema descrito nesta pesquisa.

Na Seção D.1, são transcritas as perguntas componentes do questionário supra-mencionado. Em seguida, na Seção D.2, são apresentados sumários da análise realizada sobre os dados obtidos pelos questionários. Por fim, na Seção D.3 são apresentadas as considerações finais a respeito do questionário sobre os hábitos fotográficos dos participantes.

D.1 Transcrição dos questionários realizados no experimento

O questionário sobre o perfil dos participantes foi dividido em duas partes: Na primeira parte, foram sondados dados pessoais do participante, enquanto, na segunda parte, questionaram-se hábitos fotográficos do participante.

Enunciado da questão 1: *Para a realização deste experimento, pede-se alguns dados pessoais do participante. Estes dados são para fins estatísticos e NÃO serão compartilhados NEM utilizados para fins de mala-direta.*

1. **Idade.** *Forneceu-se um campo para que o participante digite a sua idade.*
2. **Cidade/País.** *Forneceu-se um campo para que o participante digite alguma informação de localidade.*
3. **Sexo.** *Forneceu-se as opções masculino e feminino para que o participante selecionasse.*

As informações constantes no item 1 do formulário tinham apenas a finalidade de identificar se existiam padrões de comportamento fotográfico relativos à idade, região ou sexo.

Enunciado da questão 2: Antes de iniciar o experimento, é muito importante que você responda a este breve questionário sobre seus hábitos fotográficos.

Enunciado do item 1: *Sobre seus conhecimentos de fotografia, marque o que for VERDADE:*

Neste item, os participantes poderiam marcar todas as opções que considerassem pertinentes dentre as oito afirmações a seguir:

1. Você já fez algum curso sobre fotografia.
2. Você já leu algum material técnico (livros ou revistas especializadas) sobre fotografia.
3. Você gosta de suas fotografias.
4. Você ganha ou já ganhou dinheiro com suas fotografias.
5. Você recebe frequentemente elogios por suas fotos.
6. Você possui câmera fotográfica digital.
7. Após usar a sua câmera digital (caso tenha uma), você faz uma seleção das fotos apagando as piores.
8. Você prefere visualizar as fotografias em um álbum do que no computador.

Enunciado do item 2: *Com relação aos seus hábitos fotográficos, escolha a resposta mais adequada:*

Neste item, os participantes precisavam escolher apenas uma dentre as opções disponíveis:

1. Você rotineiramente faz mais de 100 fotos por mês.
2. Você rotineiramente faz até 100 fotos por mês.
3. Você apenas faz fotos em eventos isolados (viagens, festas, etc.).

Enunciado do item 3: *Estime o número de fotografias digitais que você possui em seu computador:*

Neste item, os participantes precisavam escolher apenas uma dentre as opções:

1. Menos de 10 fotos.
2. Entre 10 e 1.000 fotos.
3. Entre 1.000 e 10.000 fotos.
4. Mais de 10.000 fotos.

Enunciado do item 4: *Caso você tenha uma câmera fotográfica digital, informe seus hábitos de impressão destas fotografias em lojas especializadas:*

Neste item, os participantes precisavam escolher apenas uma dentre as opções:

1. Raramente/Nunca imprimo minhas fotos, pois prefiro visualizá-las no computador.
2. Raramente/Nunca imprimo minhas fotos pois é caro.
3. Raramente/Nunca imprimo minhas fotos pois é muito trabalhoso selecionar as que quero revelar.
4. Revelo até 30 fotografias por mês.
5. Revelo mais de 30 fotografias por mês.

O questionário 2 foi desenvolvido com o objetivo de delinear o perfil fotográfico do participante. Tal perfil pode ser utilizado com o intuito de identificar possíveis oportunidades de negócio a partir do produto proposto nesta pesquisa e qualificar, quando necessário, os participantes, e.g., podem-se selecionar apenas as fotografias dos participantes que rotineiramente obtêm mais de 10.000 fotografias por mês.

Após a exibição dos dois questionários, foram apresentadas aos participantes as instruções do experimento, conforme transcritas a seguir:

Enunciado da questão 3: *O objetivo deste experimento é identificar as principais cores de uma imagem.*

Para cada imagem exibida, você deve escolher até sete (7) cores que você considera mais representativas para a imagem - ordenadas da mais representativa para a menos representativa. Portanto, analise bem antes de iniciar a seleção das cores.

Você pode concluir o experimento em até uma semana. Para isso, clique em Continuar mais tarde e copie o link que dará acesso a sua página de votação. Para o correto funcionamento do site, não apague os cookies de seu navegador. Com quantas imagens você poderá contribuir para o experimento?

Antes de continuar, o usuário precisava selecionar o número de fotografias com as quais desejava contribuir para o experimento. As opções variaram entre 10 e 100 fotografias, tendo sido considerado 60 como o valor padrão.

Por fim, foi exibido, para cada uma das n imagens selecionadas (em que n é o número de fotografias escolhidas pelo participante no questionário 3), o questionário sobre as cores de uma imagem, o qual é apresentado na Figura D.1.

Figura D.1: Captura da tela de participação do experimento sobre cores.

A captura de tela mostra uma interface web para um experimento. No topo, há uma barra de navegação com links: 'Seja bem-vindo', 'Imagens Restantes: 1/34', 'Mais informações', 'Instruções para o experimento', 'Continuar mais tarde' e 'Sair'. O título principal é 'Identificador da Paleta de Cores'. Abaixo do título, há um texto explicativo: 'Por favor, selecione cores da imagem que você acredite representar bem o espaço de cor desta imagem. Passe o cursor do mouse sobre a imagem e clique sobre as cores que considerar relevantes. As cores devem ser selecionadas de acordo com a ordem de importância na fotografia. Você pode escolher até 7 cores.' À esquerda, há uma imagem de uma mulher com fones de ouvido. À direita, há uma paleta de cores com sete quadrados vazios, o primeiro dos quais está preenchido com uma cor azul. Abaixo da paleta, há dois botões: 'Desfazer' e 'Limpar'. Abaixo dos botões, há uma pergunta: 'Qual nota você daria a esta fotografia com relação ao seu aspecto visual?' com opções de resposta: '0 < 2.0', '3 < 4.0', '5 < 6.0', '7 < 8.0' e '9 < 10'. Abaixo da pergunta, há uma pergunta: 'Marque se você acredita necessitar de mais de 7 cores para representar esta fotografia.' com um botão 'Enviar'.

Enunciado da questão 4: *Identificador da Paleta de Cores - Por favor, selecione cores da imagem que você acredite representar bem o espaço de cor desta imagem. Passe o cursor do mouse sobre a imagem e clique sobre as cores que considerar relevantes.*

As cores devem ser selecionadas de acordo com a ordem de importância na fotografia. Você pode escolher até 7 cores.

Para realizar a seleção das cores, o participante precisou mover o cursor do mouse sobre a fotografia. No momento em que este movimento ocorria, a cor no espaço de votação era alterada para aquela apontada pelo cursor do *mouse*. No exemplo ilustrado na Figura D.1, o cursor aponta uma posição na testa do alvo humano. Tal tonalidade é então utilizada para colorir o interior da primeira caixa de resposta do lado direito. Uma vez que o participante estivesse satisfeito com a tonalidade escolhida, bastava pressionar o botão esquerdo do *mouse*. Feito isto, o quadrado em questão não modificaria mais a cor e o segundo quadrado passaria a exibir a tonalidade sob o cursor do *mouse*. Durante este processo de seleção, apenas as cores pertencentes às imagens poderiam ser selecionadas.

No experimento proposto, o participante precisou escolher no mínimo uma cor e no máximo sete cores. Caso cometesse algum erro, o participante poderia selecionar a opção “Desfazer” e, caso desejasse recomeçar o processo de seleção, poderia selecionar a opção “Limpar”.

Enunciado da questão 5: *Qual nota você daria a esta fotografia com relação ao seu aspecto visual?*

Para esta questão, o participante escolhia, dentre as alternativas disponíveis, aquela que ele considerava a mais apropriada (sendo as fotografias com qualidade inferior representadas por valores mais baixos):

1. $0 < 2,9$
2. $3 < 4,9$
3. $5 < 6,9$
4. $7 < 8,9$
5. $9 < 10$

Enunciado da questão 6: *Marque se você acredita necessitar de mais de 7 cores para representar esta fotografia.*

Para esta questão, o participante deveria apenas marcar se concordava ou não com a afirmação. Tal questão foi utilizada com o intuito de decidir se o número de cores exibidas na questão 3 era adequado para a imagem.

As questões 4, 5 e 6 foram exibidas para cada imagem do conjunto, portanto, 100 vezes.

D.2 Sumário da análise realizada sobre os dados resultantes do experimento

Uma vez concluída a etapa de rotulagem, foi realizada uma análise sobre os resultados obtidos, a qual é descrita nesta seção.

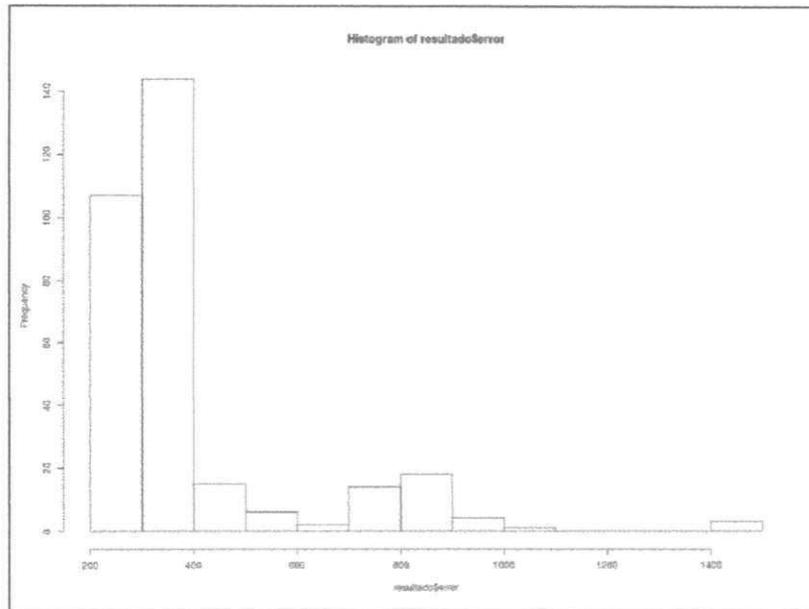
A primeira análise realizada foi a distribuição dos votantes em relação ao conjunto de imagens. Ao todo, cadastraram-se para participar do experimento 378 participantes, a partir dos quais foram geradas 8.409 rotulagens. Para evitar rotulagens tendenciosas ou incorretas (e.g., por erro do navegador), foi realizada uma filtragem, removendo-se os votos que destoaram dos demais. Dessa forma, calculou-se a diferença média entre as rotulagens de um usuário para as rotulagens dos demais. O erro E é definido pela Equação D.1.

$$E = \frac{1}{378} \times \sum_{p=1}^{378} \sum_{i=1}^{100} \sum_{c=1}^7 d(p, i, c), \quad (\text{D.1})$$

em que $d(p, i, c)$ é a distância euclidiana entre os componentes r , g e b da cor de índice c da imagem i do participante p e os componentes r , g e b a cor de índice c da imagem i do usuário sendo avaliado.

Foram removidos, do conjunto final de avaliação, aqueles usuários cujos erros superaram o valor 700. Tal limite foi determinado heurísticamente, por intermédio da avaliação dos usuários que sabidamente apresentaram erros de renderização, a partir dos quais, se pôde verificar que as rotulagens realizadas apresentaram erros superiores a 900. Adicionalmente, foi analisada a distribuição dos erros de todos os usuários e percebeu-se um comportamento irregular a partir da marca de 700, conforme apresentado na Figura D.2.

Figura D.2: Distribuição dos erros de rotulagem dos usuários



O limiar escolhido, 700, foi definido com o intuito de permitir que houvesse um alto grau de divergências entre os usuários. Também foram removidos os resultados dos usuários que participaram com vistas à calibração do experimento.

Uma vez realizada a filtragem, restaram 274 usuários e 6.624 rotulagens. Uma vez que não foi exigido que o participante rotulasse todas as 100 fotografias, muitos participantes optaram por rotular menos da metade das imagens. Para que houvesse equilíbrio na quantidade de participantes por imagem, a seleção das fotografias ocorreu aleatoriamente no subconjunto das 10 imagens menos rotuladas. A distribuição de rotulagens por imagem é ilustrada na Figura D.3.

Pode-se perceber que houve concentração na imagem cujo identificador é 0, que foi rotulada 83 vezes. Isso se deu devido a um curto período de tempo de realização do experimento em que todos os votantes iniciavam rotulando a mesma imagem. O número médio de rotulagens por imagem foi 66,24, com desvio padrão de 3,09. A imagem que menos recebeu rotulagens foi aquela cujo identificador é 98, a qual recebeu apenas 51 rotulagens.

Ao todo, 26 participantes rotularam todas as 100 imagens. A Figura D.4 contém o histograma de participação dos usuários em termos do número de imagens rotuladas.

Figura D.3: Distribuição do número de rotulagens por imagem.

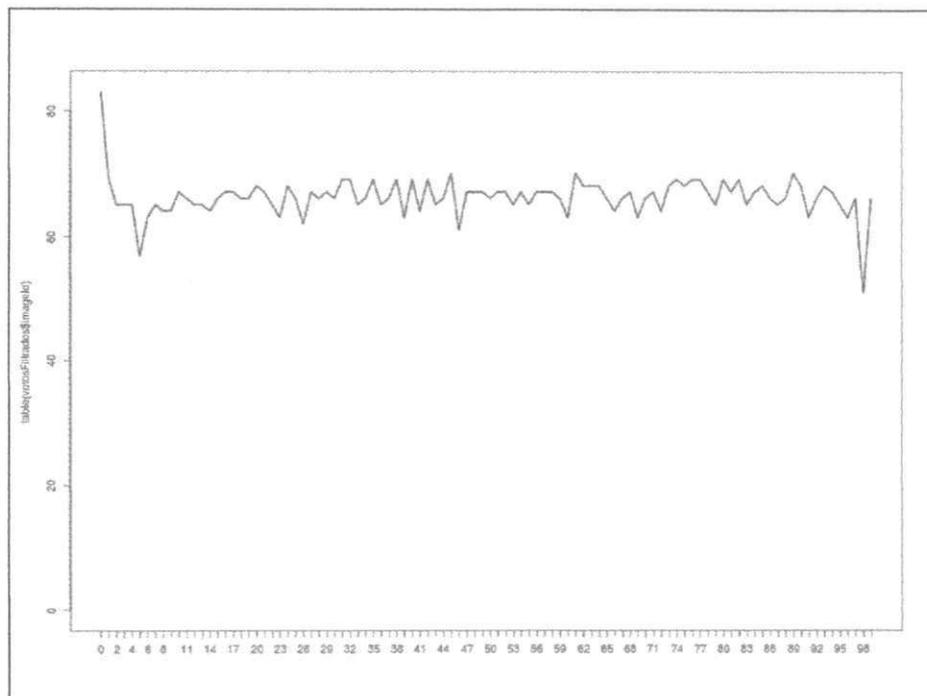
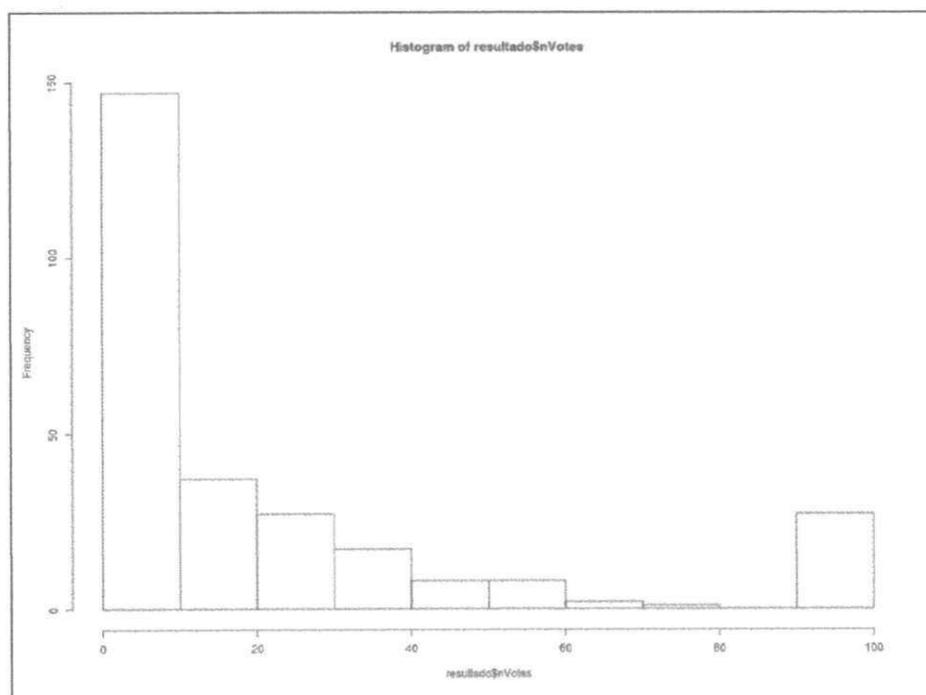


Figura D.4: Histograma do número de imagens pelo número de participantes que rotularam tal quantidade.



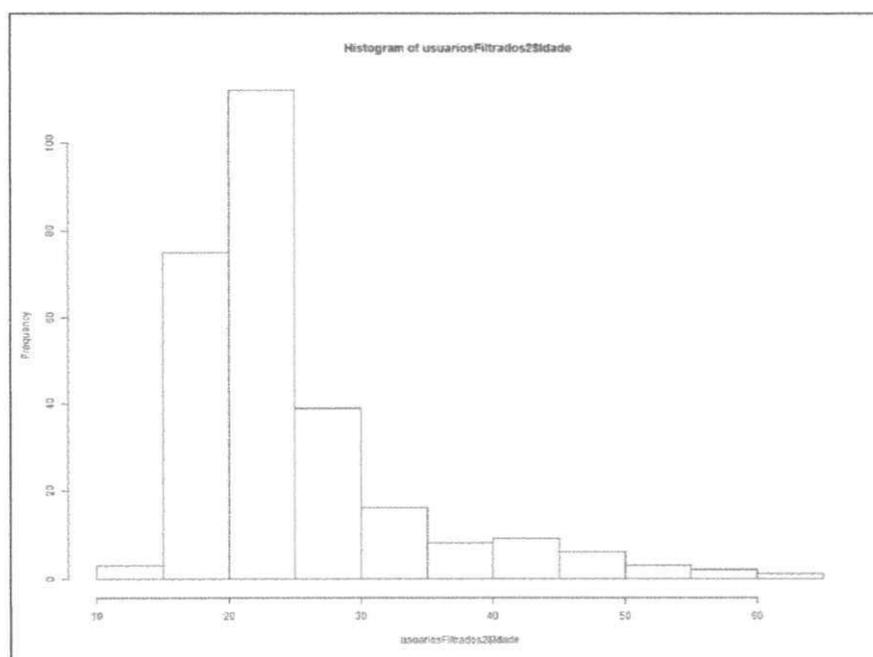
Ao analisar a Figura D.4, pôde-se perceber que à medida que o número de imagens aumenta, menos participantes rotularam, com exceção da totalidade das imagens, as quais 26 participantes rotularam. Vale salientar, adicionalmente, que nenhum usuário rotulou a mesma imagem mais de uma vez.

D.2.1 Análise do perfil dos participantes

Conforme descrito anteriormente, 274 participantes do experimento foram mantidos. Realizou-se uma análise com relação ao perfil destes participantes, a partir do questionário aplicado.

A primeira análise realizada, portanto, foi com relação à idade dos participantes. A Figura D.5 contém a distribuição da idade dos participantes.

Figura D.5: Distribuição da idade dos participantes do experimento



Pode-se concluir que grande parte dos participantes apresentava idade entre 15 e 30 anos. Quando considerados somente os participantes que concluíram as 100 imagens, o perfil se manteve.

Com relação ao gênero dos participantes, 169 participantes (62%) declararam-se do sexo masculino, enquanto 105 (38%) declararam-se do sexo feminino.

Ao analisar os hábitos fotográficos dos participantes, percebeu-se que a grande maioria, 164 participantes, ou 60% do total, obtinham fotografias apenas em eventos isolados, enquanto 60 participantes (22%) obtinham rotineiramente menos de 100 fotos em um mês e 50 participantes (18%) obtinham, rotineiramente, mais de 100 fotografias em um mês.

Os hábitos de revelação, por sua vez, permitiram inferir um perfil de participante que preferia ter as fotos apenas no computador: 97 participantes (35%) preferiam visualizá-las no computador, 67 participantes (24%) não costumavam imprimir suas fotos devido ao preço da impressão, 65 (24%) consideraram trabalhoso o ato de selecionar as fotografias. Ao mesmo tempo, para 45 participantes, a impressão de fotografias era rotineira, sendo que 38 participantes (14%) imprimiam até 30 fotografias por mês e 7 participantes (3%) imprimiam mais do que 30 fotografias.

Na última pergunta analisada, os usuários poderiam marcar quantas opções desejassem. De modo a melhor ilustrar estes resultados, as quantidades e percentuais foram exibidos na Tabela D.1.

Tabela D.1: Perfil fotográfico do participante.

Você fez algum curso sobre fotografia.	46	17%
Você já leu algum material técnico sobre fotografia.	115	42%
Você gosta de suas fotografias.	197	72%
Você ganha ou já ganhou dinheiro com suas fotografias.	35	13%
Você recebe frequentemente elogios por suas fotografias.	126	46%
Você possui câmera fotográfica digital.	244	89%
Após usar a sua câmera digital, você faz uma seleção das fotos apagando as piores.	201	73%
Você prefere visualizar as fotografias em um álbum do que no computador.	83	30%
Nenhuma das opções.	0	0%

De acordo com o perfil dos participantes, pode-se concluir que foram participantes jovens, que não possuíam um hábito fotográfico frequente e que consideraram os custos de revelação altos. Grande parte, entretanto, possuía dispositivos de captura digitais e que costumava realizar seleções em suas fotografias após a captura, descartando as piores fotografias.

Há um indicativo, portanto, de que um sistema que permita acelerar o processo de seleção de fotografias pode ter um impacto positivo neste público-alvo, fortalecendo as conclusões desta pesquisa.

D.2.2 Análise das notas atribuídas às imagens

Uma vez que também foi solicitada a opinião do participante quanto ao aspecto geral da imagem, também foi realizada uma análise das notas atribuídas às imagens. O intuito foi identificar se existia alguma correlação entre a qualidade atribuída às imagens e as correspondentes quantidades cores.

Assim, foram agrupadas as 10 piores e as 10 melhores imagens, de acordo com a votação. A Tabela D.2 contém, em ordem crescente, os identificadores, médias e desvio padrão das 10 piores fotografias. As imagens podem ser conferidas, a partir dos identificadores, no Apêndice E.

Tabela D.2: Identificador, média e desvio padrão das rotulagens das piores fotografias. As notas variam entre 0 e 5.

Identificador	Média	Desvio padrão
27	1,69	0,84
77	2,15	1,02
89	2,46	0,97
61	2,83	1,08
37	2,89	1,11
95	2,95	1,05
75	2,97	1,16
3	2,98	1,05
39	3,03	1,01
84	3,04	1,06

Em muitas das imagens existiam elementos de distração além de problemas diversos possivelmente causados pela fotogenia assim como pela qualidade da obtenção da fotografia e do armazenamento (e.g., artefatos de compressão).

Por sua vez, as médias e desvios padrão das melhores imagens são apresentadas, ordenadas em ordem crescente pela média, na Tabela D.3.

Tabela D.3: Identificador, média e desvio padrão das rotulagens das melhores fotografias. As notas variam entre 0 e 5.

Identificador	Média	Desvio padrão
17	4,07	1,00
97	4,10	0,91
64	4,13	1,09
79	4,14	0,99
18	4,15	0,86
86	4,15	0,81
32	4,17	0,86
4	4,22	0,86
46	4,31	0,85
19	4,52	0,64

Pode-se perceber, ao analisar estas imagens, que a simplicidade visual teve grande importância no julgamento realizado pelos participantes do experimento, recebendo as melhores notas as imagens que apresentaram maior grau de simplicidade.

D.3 Considerações finais

Neste apêndice, foram apresentados os questionários e seus respectivos resultados que foram aplicados aos participantes do experimento proposto no Capítulo 7.

A partir dos resultados ora apresentados, foi possível delinear (1) o perfil dos participantes, (2) as características das imagens utilizadas nos experimentos.

Apêndice E

Fotografias utilizadas no experimento de rotulagem de cores

De modo a permitir a realização do experimento descrito na Seção 7.2.2, foi necessária a construção de uma base de fotografias que pudesse ser utilizada para a obtenção da rotulagem. As fotografias pertencentes a esta base deveriam (1) ser de domínio público, (2) ter diversidade quanto aos temas, (3) apresentar variações quanto às cores utilizadas.

As Figuras E.1 a E.7 contêm as fotografias utilizadas nos experimentos. Tais fotografias foram obtidas a partir de bases públicas de imagens ou disponíveis para fins acadêmicos, as quais foram descritas na Tabela 4.1.

Figura E.1: Imagens utilizadas no experimento (1/7).



Figura E.2: Imagens utilizadas no experimento (2/7).

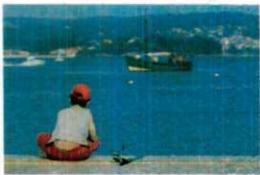
		
03	04	05
		
06	07	08
		
09	10	11
		
12	13	14
		
15	16	17
		
18	19	20

Figura E.3: Imagens utilizadas no experimento (3/7).

		
21	22	23
		
24	25	26
		
27	28	29
		
30	31	32
		
33	34	35
		
36	37	38

Figura E.4: Imagens utilizadas no experimento (4/7).

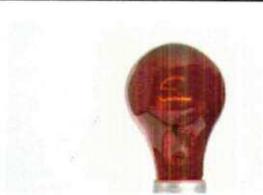
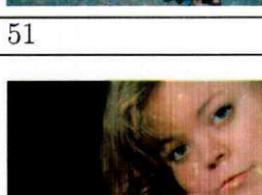
		
39	40	41
		
42	43	44
		
45	46	47
		
48	49	50
		
51	52	53
		
54	55	56

Figura E.5: Imagens utilizadas no experimento (5/7).

		
57	58	59
		
60	61	62
		
63	64	65
		
66	67	68
		
69	70	71
		
72	73	74

Figura E.6: Imagens utilizadas no experimento (6/7).

		
75	76	77
		
78	79	80
		
81	82	83
		
84	85	86
		
87	88	89
		
90	91	92

Figura E.7: Imagens utilizadas no experimento (7/7).

		
93	94	95
		
96	97	98
		
	99	

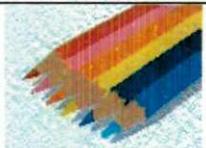
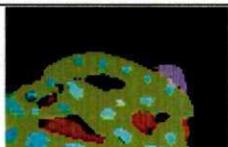
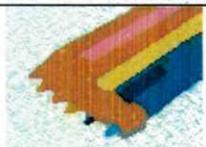
Apêndice F

Fotografias utilizadas no treinamento do segmentador

Neste apêndice, são apresentadas as imagens utilizadas no experimento relatado na Seção 7.2.1, as quais, por questões de organização e espaço, não foram ali expostas.

Nas Figuras F.1, F.2 e F.3, são apresentadas imagens utilizadas no experimento. Na primeira linha e nas duas linhas subsequentes, observam-se a imagem de referência e a imagem gerada pelo sistema após o treinamento.

Figura F.1: Exemplos de imagens da base de treinamento, imagens de referência correspondentes e imagens geradas pelo sistema treinado.

Originais				
Referência				
Geradas				

Uma vez que o objetivo deste método é a verificação da harmonia das cores de uma imagem, foram extraídas as três principais cores de cada uma das imagens processadas.

Figura F.2: Exemplos de imagens da base de treinamento, imagens de referência correspondentes e imagens geradas pelo sistema treinado.

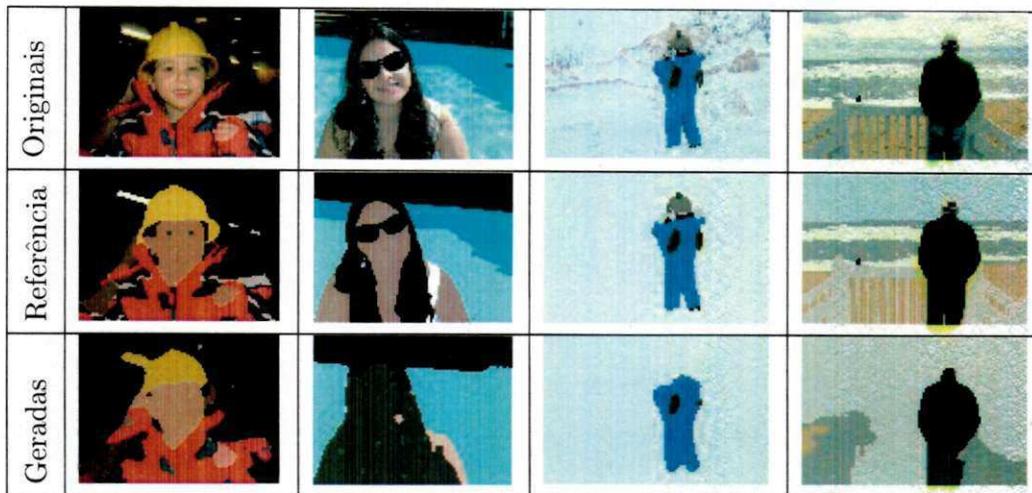
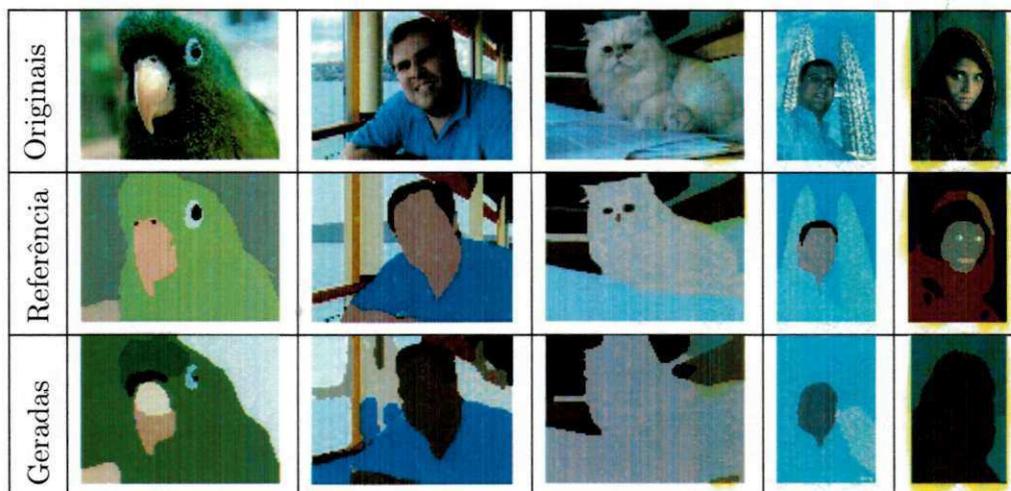


Figura F.3: Exemplos de imagens da base de treinamento, imagens de referência correspondentes e imagens geradas pelo sistema treinado.



As cores foram obtidas a partir da quantidade de pontos das regiões segmentadas. As cores extraídas pelo método proposto são posicionadas lado a lado com as cores extraídas pelo método OSCAR (YAO et al., 2011), a partir de uma aplicação de *software* demonstrativa *online*¹². Assim, nas Figuras F.4 a F.12 são apresentados os resultados da extração das cores principais das imagens.

Pode-se perceber, ao analisar os resultados, que os métodos, de certa forma, se equiparam. Uma vez que o objetivo é extrair as três principais cores que compõem a

¹²<http://acquine.alipr.com/oscar/>

imagem (na ordem de exibição), pode-se afirmar que ambos os métodos cumpriram com o objetivo proposto. Entretanto, em alguns casos, houve inconsistência nos resultados. Na imagem 8 (Figura F.7), por exemplo, o método OSCAR não levou em consideração a tonalidade do plano de fundo, tendo o mesmo ocorrido com a imagem 10 (Figura F.8). Por outro lado, na imagem 12 (Figura F.9), o método proposto não selecionou, dentre as principais cores, tons importantes (e.g., tom da areia e da camisa do alvo). Espera-se a melhoria dos resultados após o aprimoramento do método de segmentação e com diferentes estratégias para a seleção das tonalidades principais da imagem.

Figura F.4: Exemplos de imagens produzidas e cores extraídas.

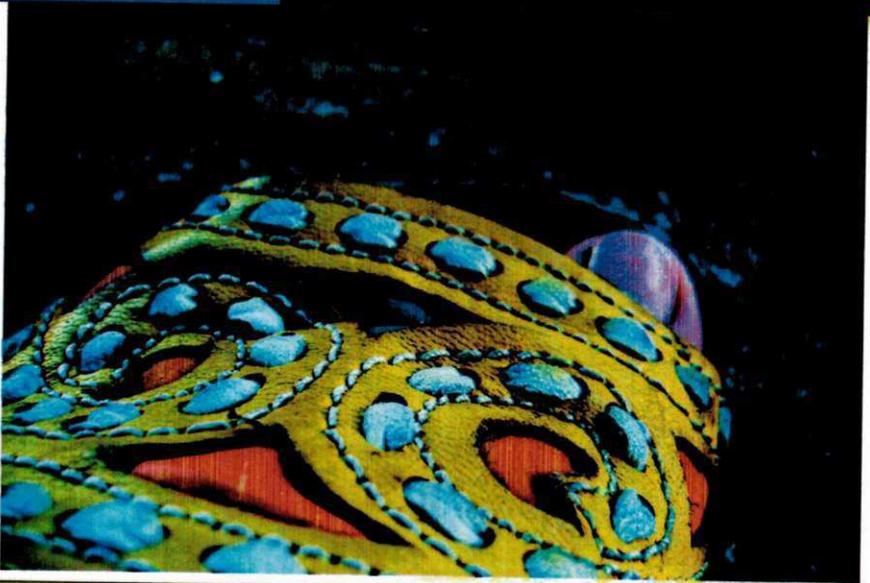
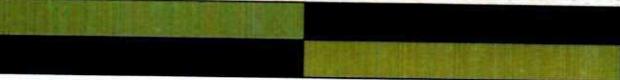
Imagem 1	
Proposto	
OSCAR	
Imagem 2	
Proposto	
OSCAR	

Figura F.5: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.

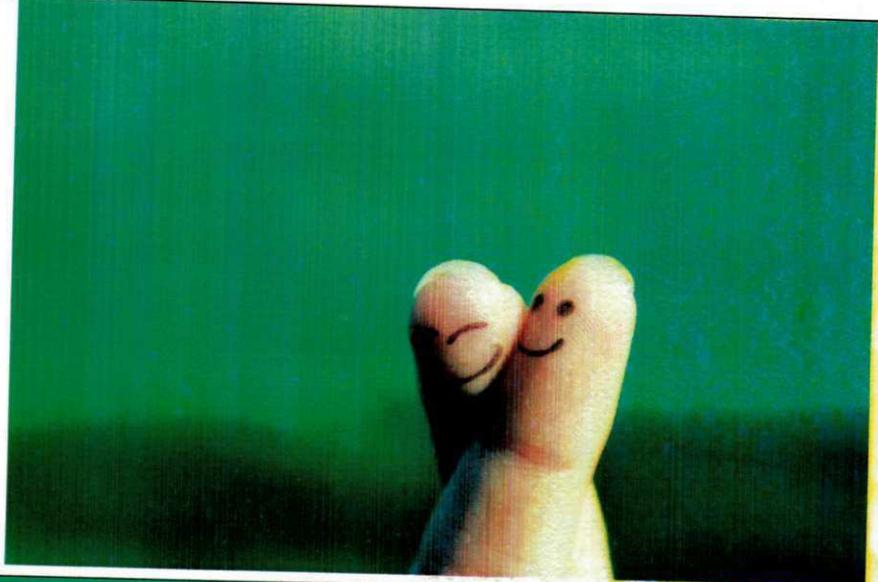
Imagem 3	
Proposto	
OSCAR	
Imagem 4	
Proposto	
OSCAR	

Figura F.6: Exemplos de imagens produzidas e cores extraídas.

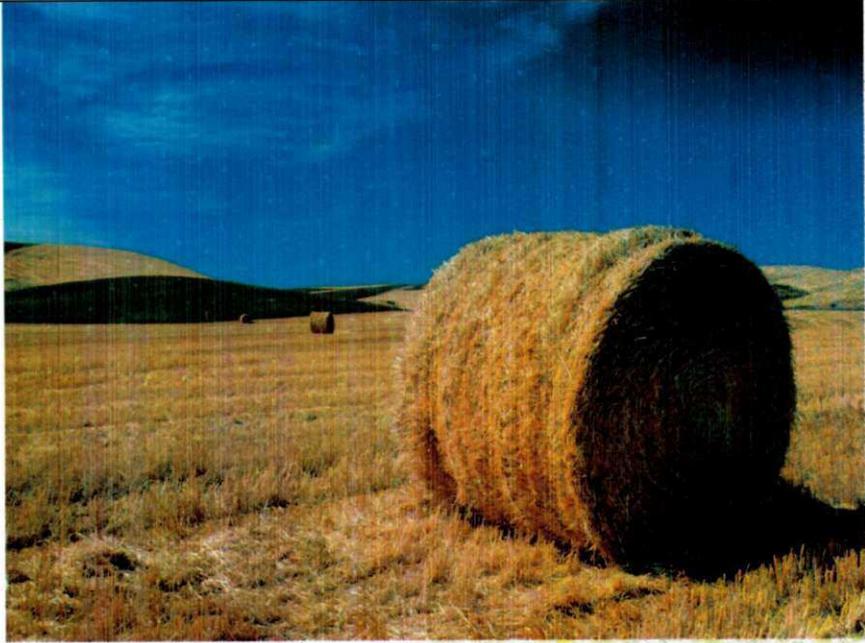
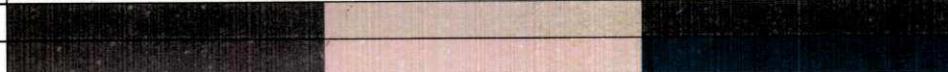
<p>Imagem 5</p>	
<p>Proposto</p>	
<p>OSCAR</p>	
<p>Imagem 6</p>	
<p>Proposto</p>	
<p>OSCAR</p>	

Figura F.7: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.

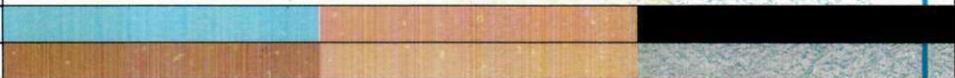
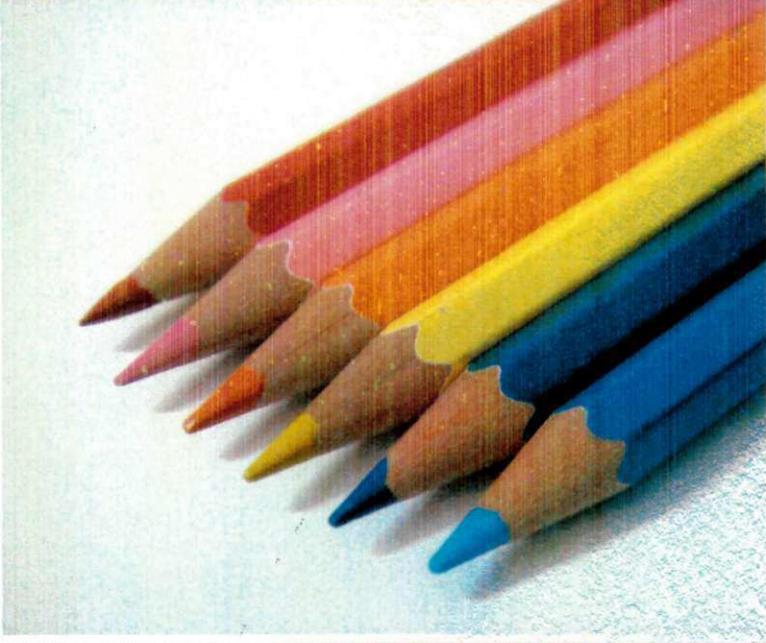
Imagem 7	
Proposto	
OSCAR	
Imagem 8	
Proposto	
OSCAR	

Figura F.8: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.

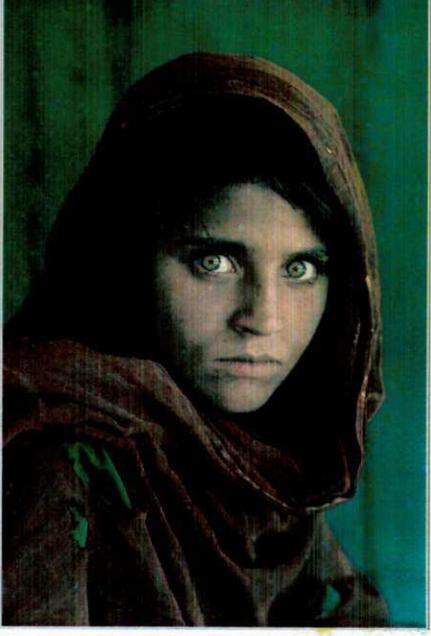
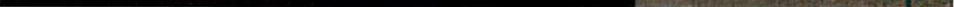
Imagem 9	
Proposto	
OSCAR	
Imagem 10	
Proposto	
OSCAR	

Figura F.9: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.

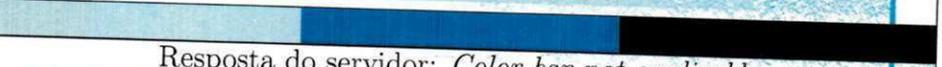
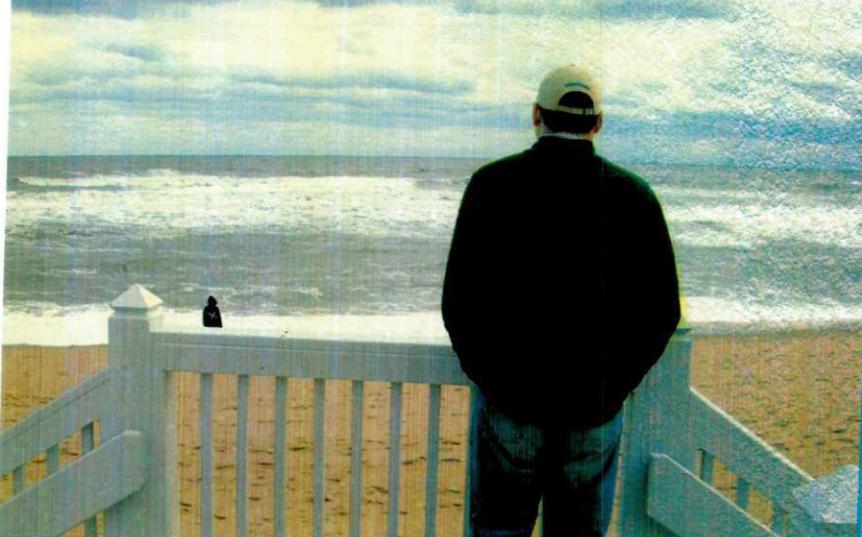
Imagem 11	
Proposto	
OSCAR	Resposta do servidor: <i>Color bar not applicable</i>
Imagem 12	
Proposto	
OSCAR	

Figura F.10: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.

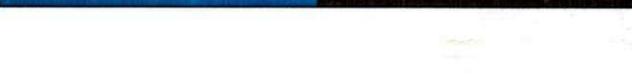
Imagem 13	
Proposto	
OSCAR	
Imagem 14	
Proposto	
OSCAR	

Figura F.11: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.

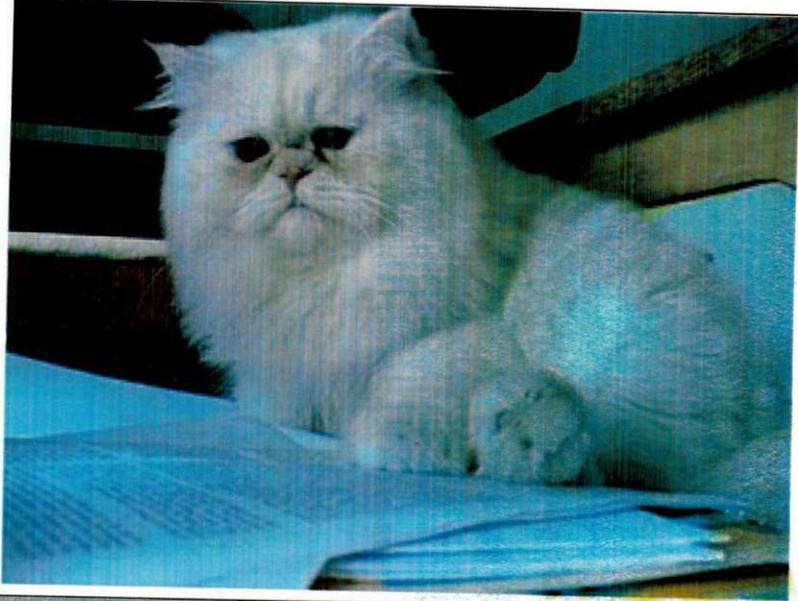
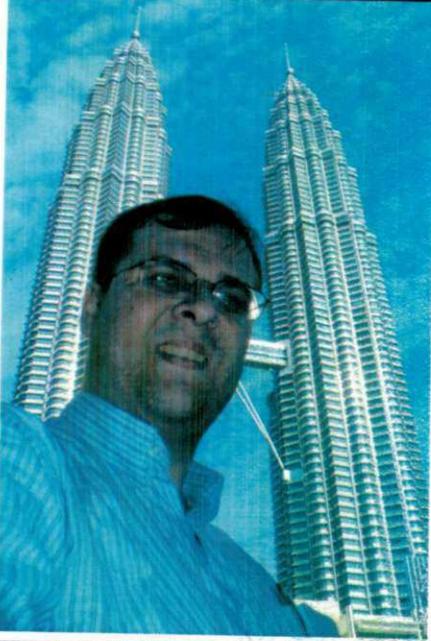
Imagem 15	
Proposto	
OSCAR	
Imagem 16	
Proposto	
OSCAR	

Figura F.12: Exemplos de imagens e as cores extraídas utilizando o método proposto e o método OSCAR.



Apêndice G

Diário do evento

O experimento subjetivo, descrito na Seção 7.4, necessitava que os participantes selecionassem as fotografias preferidas pertencentes a um mesmo evento. O objetivo desta seleção era reproduzir, com o máximo de fidelidade, a situação na qual um indivíduo escolhia, de acordo com sua opinião, as fotografias que melhor representavam o evento. Entretanto, há de se destacar que existe um grande teor de subjetividade neste processo seletivo, uma vez que cada lugar (ou indivíduo) tem um significado diferente para o proprietário das fotos e um terceiro indivíduo.

De modo a reduzir a subjetividade da tarefa, foi redigido um diário contendo as fotografias e os textos descrevendo os eventos aos quais as fotografias se referem. Este diário foi apresentado a cada usuário antes do início do experimento.

O diário é composto de um texto explicativo, descrevendo um sub-evento com o intuito de contemporizar as fotografias. Para tanto, foi introduzida a história de um personagem fictício, *Severina*, e sua viagem à cidade de *Buenos Aires, Argentina*. Apesar de textos e personagens não serem reais, esta informação não foi repassada aos participantes do experimento.

Tal conjunto de fotografias foi escolhido por ter sido obtido em um intervalo de tempo adequado (nem menor que um dia nem maior que uma semana), possuir um enredo simples de ser compreendido em um curto espaço de tempo, conter fotografias de nível amador, tanto de boa como de má qualidade, não ser muito volumoso (são 144 fotografias) e possuir subeventos de duração variável.

G.1 Diário do evento

Estas são as fotos tiradas por Severina em sua última viagem para Buenos Aires. Severina pretende produzir um álbum fotográfico destas fotografias e precisa de sua ajuda para realizar a seleção das fotografias que melhor descrevem a viagem.

Severina preparou este documento para você entender melhor como foi a viagem, associando fotos à descrição de cada ocasião que ela considerou relevante.

Viagem a Buenos Aires

Esta viagem a Buenos Aires era um sonho antigo e realizá-lo com minha nora Heloísa e meu filho Adalberto o tornou ainda mais completo.

Chegamos de madrugada e fomos direto ao Hotel. No dia seguinte, ao acordar, tive uma grata surpresa: a janela do hotel dava para um dos pontos turísticos da cidade, o Obelisco.

Figura G.1: Imagens utilizadas no experimento (1/21).



O centro da cidade é belíssimo! Destaco a catedral de Buenos Aires, de onde, inclusive, saiu o novo papa.

Figura G.2: Imagens utilizadas no experimento (2/21).



No dia seguinte, fomos conhecer o delta do rio Tigre, um passeio que recomendo. Além da agradável viagem de trem, passeamos de barco pelo rio Tigre. Foi possível conhecer toda a região litorânea do rio, com suas casas e observar o tráfego marítimo do lugar.

Figura G.3: Imagens utilizadas no experimento (3/21).



Figura G.4: Imagens utilizadas no experimento (4/21).



Figura G.5: Imagens utilizadas no experimento (5/21).



Ainda no mesmo dia, mas já quase à noite, aproveitamos para conhecer a sede do governo argentino: a Casa Rosada. Sem dúvidas, um passeio histórico. Pudemos ver quadros presenteados pelo governo brasileiro, vasos antigos, o suntuoso interior do palácio e o lado externo, repleto de singulares detalhes.

Figura G.6: Imagens utilizadas no experimento (6/21).

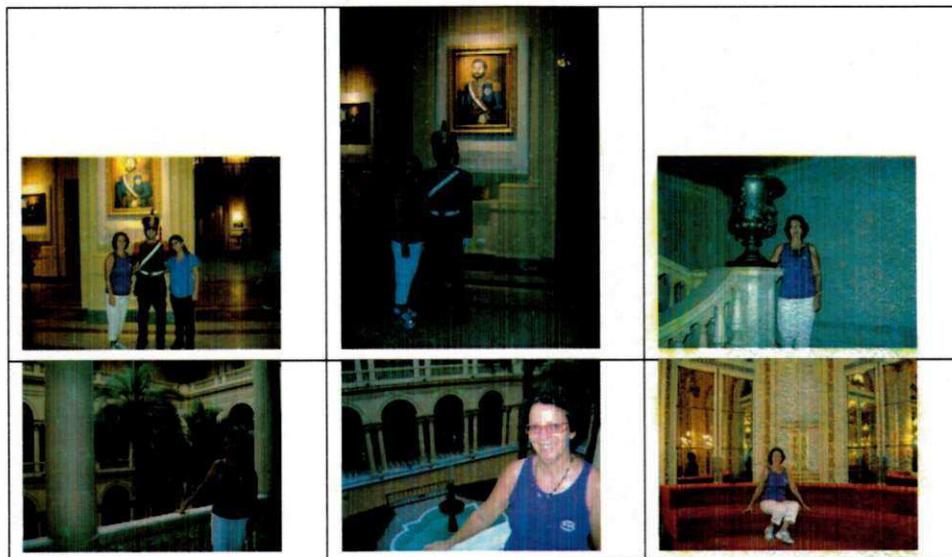
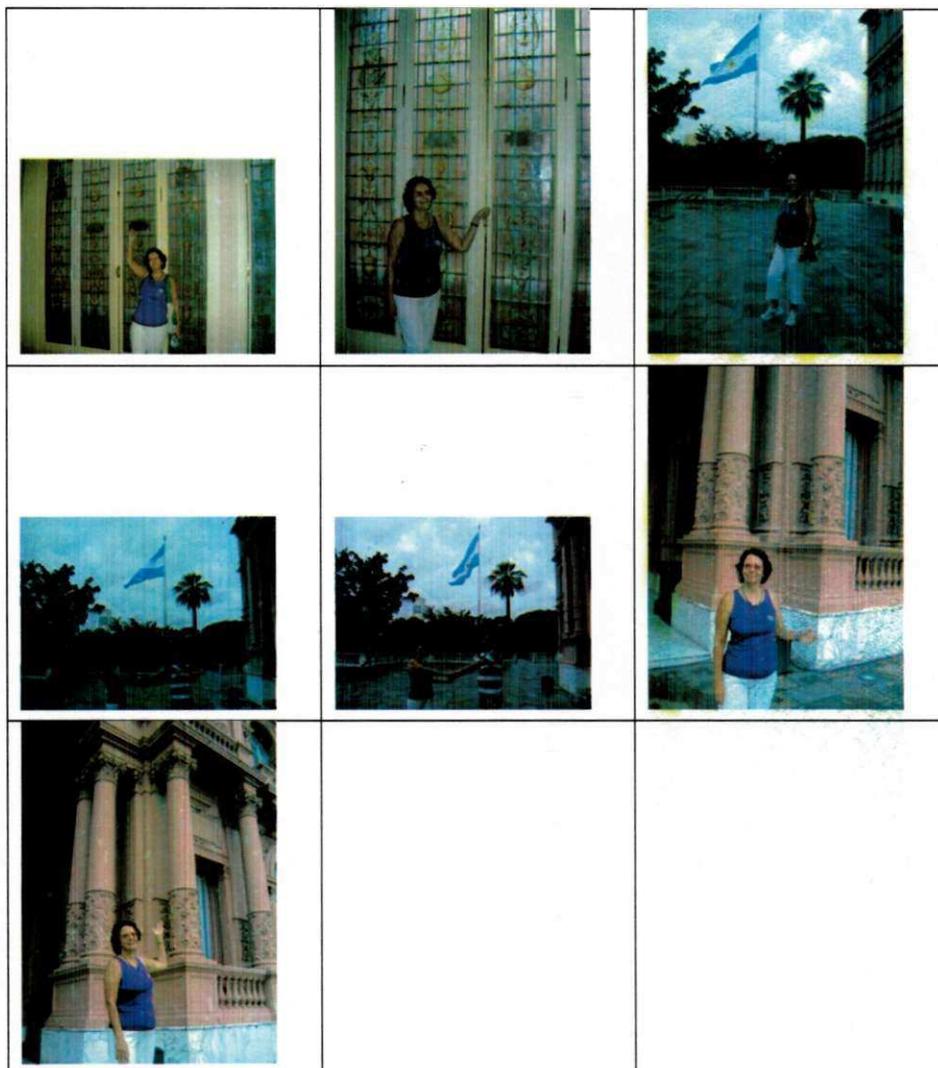


Figura G.7: Imagens utilizadas no experimento (7/21).



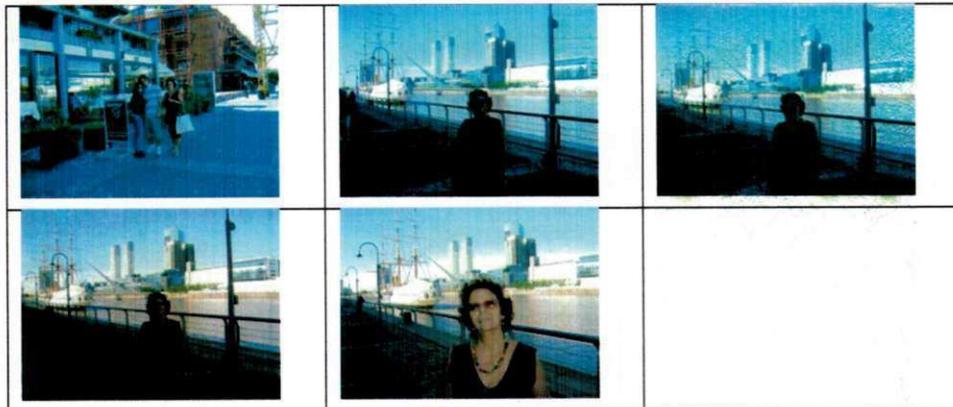
No dia seguinte, voltamos ao Tigre, dessa vez pela linha convencional de trem. Registrei este momento na bela estação terminal.

Figura G.8: Imagens utilizadas no experimento (8/21).



Na volta do Tigre, fomos ao moderno Puerto Madero, o qual é um porto, como sugere o nome, cheio de restaurantes, cassinos flutuantes e a Ponte da Mulher, outro ponto turístico da cidade.

Figura G.9: Imagens utilizadas no experimento (9/21).



À noite, fomos a uma tradicional apresentação de tango. Não pude tirar fotos durante a apresentação, só antes de sair, ainda no lobby do hotel e durante a refeição inclusa no pacote.

Figura G.10: Imagens utilizadas no experimento (10/21).



Mais um dia e mais um passeio. Dessa vez, um passeio triplo: Jardim Botânico, Jardim Japonês e o Rosedal. Primeiro, fomos ao Jardim Botânico.

Figura G.11: Imagens utilizadas no experimento (11/21).

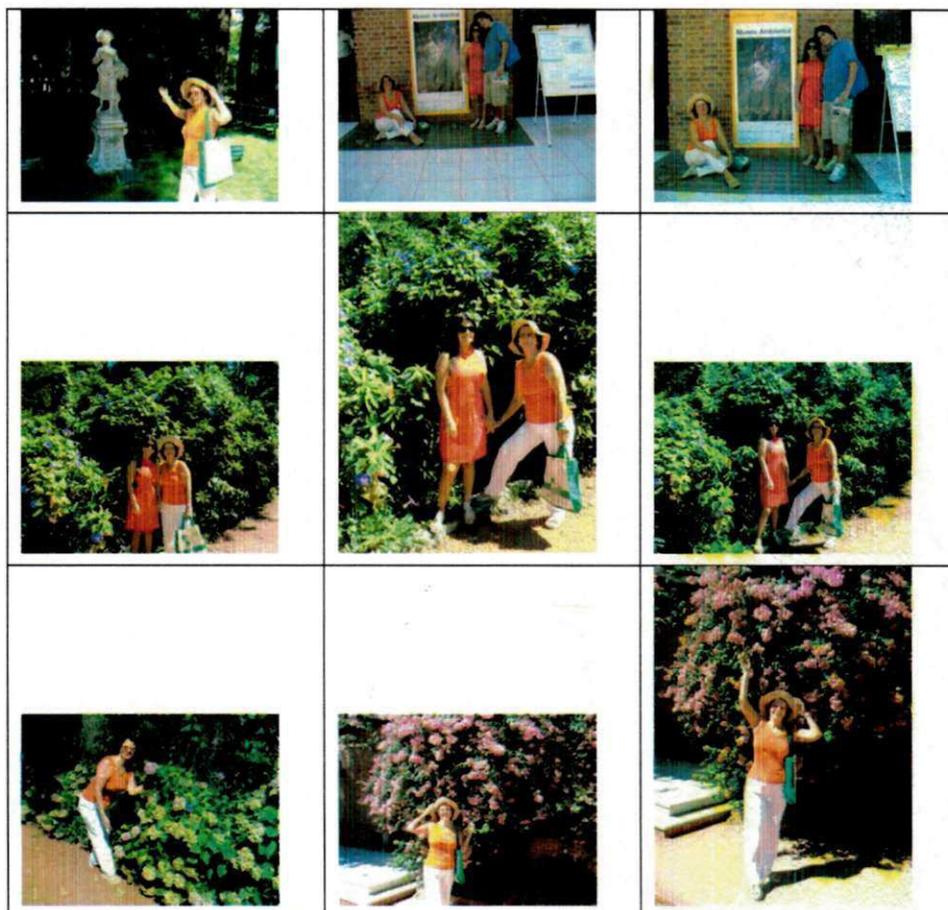


Figura G.12: Imagens utilizadas no experimento (12/21).

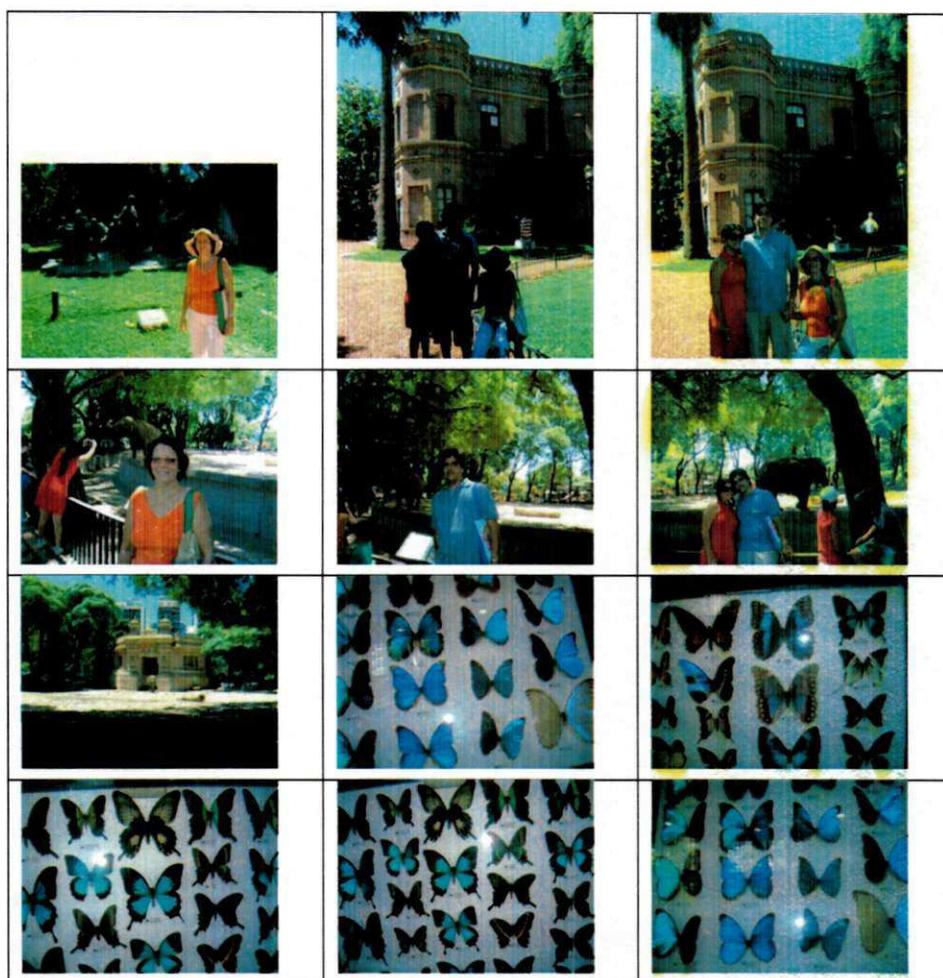
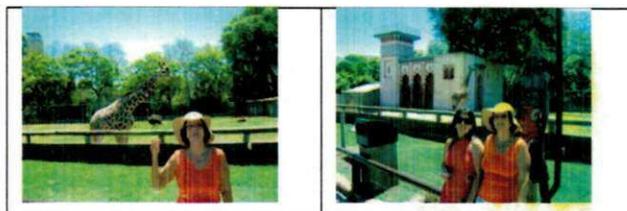


Figura G.13: Imagens utilizadas no experimento (13/21).

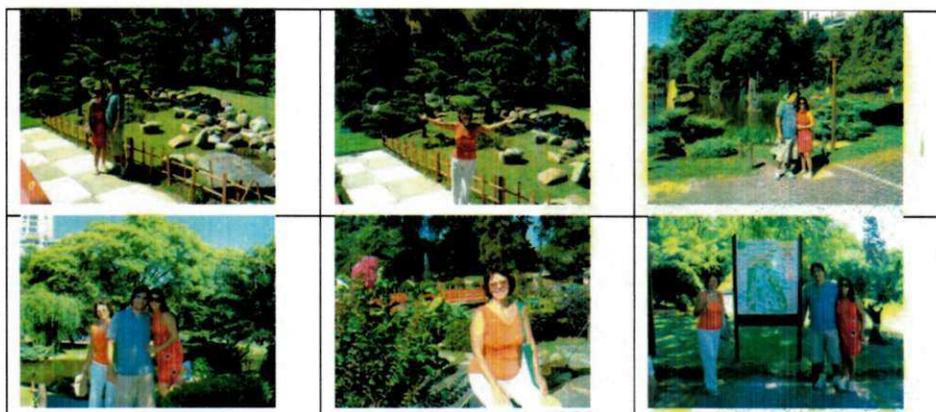


Pertinho dali estava o Jardim Japonês, um local de singular beleza.

Figura G.14: Imagens utilizadas no experimento (14/21).



Figura G.15: Imagens utilizadas no experimento (15/21).



Em seguida fomos ao Parque 3 de Fevereiro, mais conhecido como El Rosedal.

Figura G.16: Imagens utilizadas no experimento (16/21).

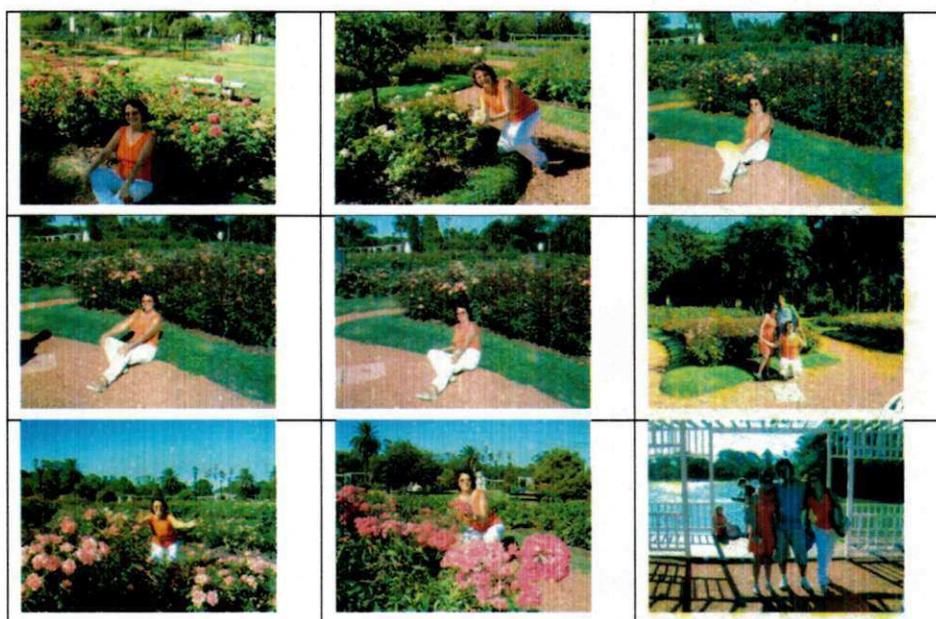
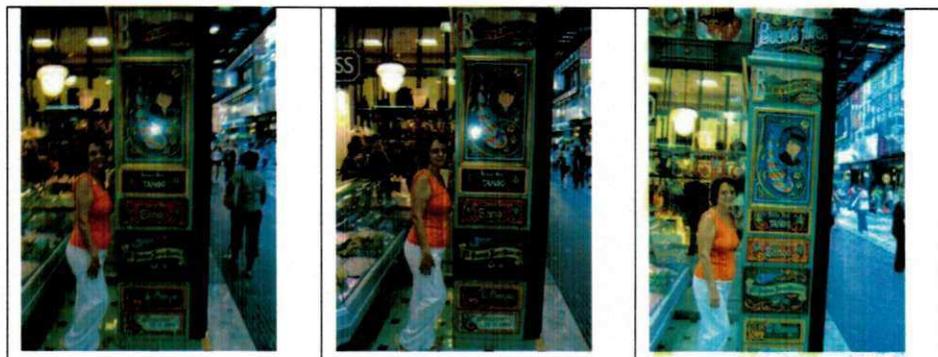


Figura G.17: Imagens utilizadas no experimento (17/21).



Na volta, fizemos uma rápida parada na Calle Florida

Figura G.18: Imagens utilizadas no experimento (18/21).



De noite, nos preparamos para jantar em um restaurante muito belo: o Las Violetas.

Figura G.19: Imagens utilizadas no experimento (19/21).



No penúltimo dia, passeamos pouco, com destaque para o trio peruano no centro da capital argentina.

Figura G.20: Imagens utilizadas no experimento (20/21).



No dia seguinte, o último passeio: Aeroporto de Ezeiza!

Figura G.21: Imagens utilizadas no experimento (21/21).



G.2 Considerações finais

Neste apêndice, foi descrito o processo de criação do diário do viajante, o qual foi apresentado aos participantes no início do experimento descrito na Seção 7.4.

O diário em questão teve como objetivo contemporizar as fotografias, permitindo aos participantes do experimento, identificar os momentos em que as fotografias foram obtidas. A noção de tempo permitiu ao participante avaliar melhor a distribuição das fotografias ao longo das páginas do álbum.