
Divergências de Bregman e Total Bregman aplicadas na análise de imagens

Daniela Portes Leal Ferreira



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Uberlândia
2017

Daniela Portes Leal Ferreira

**Divergências de Bregman e Total Bregman
aplicadas na análise de imagens**

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Celia A. Zorzo Barcelos

Uberlândia
2017

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

F383d Ferreira, Daniela Portes Leal, 1972-
2017 Divergências de Bregman e Total Bregman aplicadas na análise de
imagens [recurso eletrônico] / Daniela Portes Leal Ferreira. - 2017.

Orientadora: Celia Aparecida Zorzo Barcelos.

Tese (Doutorado) - Universidade Federal de Uberlândia, Programa
de Pós-Graduação em Ciência da Computação.

Modo de acesso: Internet.

Disponível em: <http://dx.doi.org/10.14393/ufu.te.2018.791>

Inclui bibliografia.

Inclui ilustrações.

1. Computação. 2. Processamento de imagens. 3. Algoritmos
genéticos. I. Barcelos, Celia Aparecida Zorzo, (Orient.) II. Universidade
Federal de Uberlândia. Programa de Pós-Graduação em Ciência da
Computação. III. Título.

CDU: 681.3

Maria Salete de Freitas Pinheiro - CRB6/1262

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO

Os abaixo assinados, por meio deste, certificam que leram e recomendam para a Faculdade de Computação a aceitação da Tese de doutorado intitulada **“Divergências de Bregman e Total de Bregman aplicadas na análise de imagens”** por **Daniela Portes Leal Ferreira** como parte dos requisitos exigidos para a obtenção do título de **Doutora em Ciência da Computação**.

Uberlândia, 29 de agosto de 2017

Orientadora:

Prof. Dra. Celia Aparecida Zorzo Barcelos
Universidade Federal de Uberlândia

Banca Examinadora:

Prof. Dr. João Espírito Santo Batista Neto
Universidade de São Paulo

Prof. Dr. Marcos Aurélio Batista
Universidade Federal de Goiás

Prof. Dr. Ricardo da Silva Torres
Universidade Estadual de Campinas

Prof. Dr. Bruno Augusto Nassif Travençolo
Universidade Federal de Uberlândia

Prof. Dr. Marcelo Keese Albertini
Universidade Federal de Uberlândia

Ao meu companheiro e amigo Anderson e aos meus filho, Lucas e Nathalia, bençãos de Deus na minha vida que me apoiaram, amaram, compreenderam e foram minha companhia durante toda a trajetória que me levou à concretização deste trabalho. Aos meus pais Hélia e Anízio que formaram os alicerces de minha história.

Agradecimentos

A “gratidão é a memória do coração”. Em minha trajetória sou sempre surpreendida por “anjos da guarda” enviados para me ajudar, sem os quais seria muito difícil alcançar meus objetivos. Assim quero expressar aqui, a minha gratidão a essas pessoas que sem dúvida são presentes de Deus para mim e que ao longo dos anos de trabalho, que resultaram nesta tese, me ajudaram, ensinaram e apoiaram.

Agradeço à minha professora, educadora, orientadora, e principalmente amiga, Dra Celia A. Zorzo Barcelos. Meu muito obrigada pela sua generosidade e dedicação, por acreditar no meu potencial, pela orientação e principalmente pelo bom convívio nestes anos de trabalho.

Agradeço, também, ao professor Eraldo Ribeiro pela oportunidade de trabalho e rico aprendizado, e ao professor André Backes, pela disponibilidade e apoio.

Agradeço, em especial, aos colegas, Bruno Moraes (meu irmãozinho) e Fabíola Fernandes (um dos meus anjos da guarda) pela contribuição nos trabalhos enviados e apresentados nos congressos.

Meus agradecimentos aos colegas Crícia, Ernani, Thiago, Leandro e André pela alegre convivência e pela boa disposição com que realizamos os muitos trabalhos em comum.

Deixo também uma palavra de agradecimento aos professores Bruno Travençolo, Gina Oliveira, Maria Camila Nardini e Rita Julia, pela dedicação com que ministraram suas disciplinas, enriquecendo, com certeza, a minha formação.

Finalmente, agradeço a todos que torceram ou intercederam por mim durante esta jornada.

“Senhor, tu me sondas e me conheces. Sabes quando me sento e quando me levanto; de longe percebes os meus pensamentos. Sabes muito bem quando trabalho e quando descanso; todos os meus caminhos te são bem conhecidos. Antes mesmo que a palavra me chegue à língua, tu já a conheces inteiramente, Senhor. Tu me cercas, por trás e pela frente, e põe a tua mão sobre mim. Tal conhecimento é maravilhoso demais e está além do meu alcance, é tão elevado que não o posso atingir. Tuas obras são maravilhosas! Disso tenho plena certeza.”

Resumo

As divergências de Bregman e Total Bregman são úteis para determinar a similaridade de dados complexos e têm sido utilizadas em diversas aplicações. Algoritmos fundamentais e estruturas de dados foram generalizados oferecendo, assim, meta-algoritmos que podem ser aplicados utilizando qualquer divergência de Bregman. Considerando a relevância das generalizações de métodos utilizando as divergências de Bregman e Total Bregman, visto que estas são medidas de dissimilaridade não métricas, propõe-se, neste trabalho, novos métodos de análise de imagens definidos para estas classes de divergências. Nesta perspectiva, são definidos novos funcionais de energia que possibilitam a generalização do método de segmentação hierárquica baseado no funcional de Mumford Shah e do método variacional empregado no registro de imagens. A aplicação destas divergências exigiu a defnição de condições e tratamentos adequados. Os funcionais e tratamentos apresentados foram empregados na análise de imagens reais e sintéticas. Os resultados obtidos demonstram a viabilidade da aplicação dos funcionais propostos e evidenciam que a definição de tratamentos considerando as características dos dados e os domínios de aplicacão das divergências propiciam um melhor desempenho dos métodos aplicados na análise de imagens.

Palavras-chave: Divergências de Bregman, Divergências Total Bregman, Similaridade, Processamento de Imagens.

Abstract

The Bregman and Total Bregman divergences are useful for determining the similarity of complex data and have been used in various applications. Fundamental algorithms and data structures have been generalizes offering thus meta-algorithms that can be applied using any Bregman divergences. Considering the relevance of generalizations methods using Bregman and Total Bregman divergences, since they are not metric dissimilarity measures, it is proposed in this work, new methods of image analysis defined for these class of Bregman divergence measures. In this perspective, we have developed new functional energy that enables the generalization of the hierarchical segmentation method based on functional Mumford Shah and the generalization of the variational method used in image registration. Conditions and treatments suitable to support similarity search defined by these divergences were established. Both the functional and the treatments were employed in the analysis of real and synthetic images. The results demonstrate the viability of implementing the defined functionals and show that the treatments, considering the characteristics and differences of application domains, provide optimization of the methods used in image analysis.

Keywords: Bregman Divergences, Total Bregman Divergences, Similarity, Image Processing.

Listas de ilustrações

Figura 1 – Conceitos principais, contribuições e publicações.	35
Figura 2 – Exemplo de Dendograma.	40
Figura 3 – Exemplo do cálculo de a, b, c e d	43
Figura 4 – Exemplo das medidas $a(i)$ e $b(i)$ para o cálculo da Silhueta.	44
Figura 5 – (a) Representação de uma segmentação hierárquica. (b) Estrutura de Árvore da partição hierárquica.	45
Figura 6 – Segmentações de uma imagem utilizando diferentes níveis de detalhes. .	46
Figura 7 – Grafo de Regiões Adjacentes (RAG).	46
Figura 8 – Localização das partições.	47
Figura 9 – Comparação das regiões de referência R_i e segmentação S_i	49
Figura 10 – Fluxograma da abordagem BoVW combinada com os métodos de atenção visual e pirâmides espaciais.	51
Figura 11 – Visão geral do <i>Bag of Visual Words</i>	52
Figura 12 – Interpretação geométrica da divergência de Bregman no \mathbb{R}^2	62
Figura 13 – Exemplo de Árvore de Bolas de Bregman	66
Figura 14 – Bolas de Bregman	66
Figura 15 – Partição do Espaço utilizando diferentes DB	67
Figura 16 – (a)Conjunto de dados, (b) Clusters obtidos utilizando a divergência <i>Squared loss</i> (c) Clusters resultantes utilizando a divergência <i>Itakura Saito</i>	67
Figura 17 – Interpretação geométrica da divergência Total Bregman.	68
Figura 18 – Diferença de comportamento entre DB e DTB: (a) <i>Squared loss</i> , (b) <i>Total Squared loss</i> , (c) <i>Generalized I-divergence</i> , (d) <i>Total Generalized I-divergence</i> , (e) <i>Logistic loss</i> , (f) <i>Total Logistic loss</i> , (g) <i>Itakura-Saito</i> (h) <i>Total Itakura-Saito</i> , (i) <i>Hellinger-like</i> and (j) <i>Total Hellinger-like</i> . .	70
Figura 19 – O impacto do parâmetro ε na função GID/ε , medida de avaliação MP@10 no banco de dados Caltech101 e abordagem BoVW.	74

Figura 20 – O impacto do parâmetro α na função GID/D, medida de avaliação MP@10m no banco de dados Caltech101 e abordagem BoVW.	74
Figura 21 – O impacto do parâmetro ε na função KL/ ε , medida de avaliação MP@10 no banco de dados Caltech101.	76
Figura 22 – O impacto do parâmetro α na função GID/D, medida de avaliação MP@10 no banco de dados Caltech101.	76
Figura 23 – Resultados da recuperação obtida aplicando os tratamentos KL/N e KL/TI	77
Figura 24 – Bolas Total Bregman	81
Figura 25 – Partição do Espaço utilizando TBB-trees	84
Figura 26 – (a)Conjunto de dados (JAIN; MARTIN, 2005), (b, c) <i>Clusters</i> obtidos utilizando a <i>kd-tree</i> e <i>k-means</i> , respectivamente, (d, e, f) <i>Clusters</i> resultantes utilizando a TBB-tree com a divergência <i>Total Squared loss</i> , <i>Total Itakura Saito</i> e <i>Total Generalized I-divergence</i> , respectivamente.	87
Figura 27 – Partição inicial do espaço utilizando TBB-trees e t-centros	87
Figura 28 – Eficácia da TBB-tree na clusterização de dados complexos	88
Figura 29 – Average Precision em 100.	90
Figura 30 – (a) Imagens Originais, (b) Imagens segmentadas utilizando <i>Squared loss</i> , (c)Imagens segmentadas utilizando <i>Generalized I-divergence</i> , (d) Imagens segmentadas utilizando <i>Itakura Saito</i>	98
Figura 31 – Planejamento dos experimentos	111
Figura 32 – BD entre os pixels de duas imagens médicas utilizando Logistic loss com (a) $\epsilon = 0.01$, (b) $\epsilon = 0.001$	113
Figura 33 – Imagens Registradas x Cor. Primeira linha: Diferença inicial. Primeira coluna: Diferenças finais obtidas utilizando as divergências: <i>Squared loss</i> , <i>Logistic loss</i> , <i>Generalized I-divergence</i> e <i>Itakura Saito</i> . Segunda coluna: Imagens finais, utilizando as divergências: <i>Squared loss</i> (S^{sl}), <i>Logistic loss</i> (S^{ll}), <i>Generalized I-divergence</i> (S^{gid}) e <i>Itakura Saito</i> (S^{is}).	114
Figura 34 – Coeficiente de correlação- “cor”x Iterações.	115
Figura 35 – Imagens Registradas com as caracterizações “cor e entropia” e “cor e Média”. (a) <i>Squared loss</i> , (b) <i>Logistic loss</i> , (c) <i>Generalized I-divergence</i> e (d) <i>Itakura Saito</i>	116
Figura 36 – Coeficiente de correlação “cor e entropia” x Iterações.	116
Figura 37 – Coeficiente de correlação- “cor e Média” x Iterações.	117
Figura 38 – Imagens Registradas com as caracterização “momentos”. (a) <i>Squared loss</i> , (b) <i>Logistic loss</i> , (c) <i>Generalized I-divergence</i>	117
Figura 39 – Imagens Registradas com a caracterização “cor, media, momentos, entropia e borda”. (a) Distância <i>Squared loss</i> , (b) divergência <i>Logistic loss</i> , (c) <i>Generalized I-divergence</i> e (d) divergência <i>Itakura Saito</i>	117

Figura 40 – Coeficiente de correlação- <i>Itakura Saito</i>	118
Figura 41 – Coeficiente de correlação- <i>Squared loss</i>	118
Figura 42 – Coeficiente de correlação- <i>Generalized I-divergence</i>	119
Figura 43 – Coeficiente de correlação - <i>Logistic loss</i>	119
Figura 44 – (a) Imagem de referência S, (b) Imagem modelo T e (c) Diferença inicial entre S e T	120
Figura 45 – Melhores resultados para as divergências <i>Squared loss</i> , <i>Generalized I-divergence</i> e <i>Logistic loss</i> . (a) imagem modelo T, (b) diferença entre a imagem T e as imagens finais, $S(x + u(x))$ e (c) imagens finais.	121
Figura 46 – Nas linhas tem-se as imagens S, T, a diferença inicial $S(x)-T(x)$, a diferença final $S^{eucl}(x + u) - T(x)$ obtida com a divergência <i>Squared loss</i> , a diferença final $S^{ll}(x+u)-T(x)$ obtida com a divergência <i>Logistic loss</i> e a diferença final $S^{id}(x + u) - T(x)$ considerando a divergência <i>Generalized I-divergence</i> para as imagens médicas: próstata, coração, peito, abdômen (corte 1) e abdômen (corte 2).	123
Figura 47 – Evolução do CC resultante do registro do par de imagens próstata considerando as divergências <i>Squared loss</i> , <i>Logistic loss</i> e <i>Generalized I-divergence</i>	124
Figura 48 – Evolução do CC em relação ao número de iterações após 5 minutos de processamento.	125
Figura 49 – (a) Evolução da variação do CC para cada par de imagem considerando as diferentes divergências. (b) Efeito do fator “Divergência” sobre o aumento relativo coeficiente de correlação das imagens registradas considerando a caracterização “cor”	125
Figura 50 – Evolução do coeficiente de correlação para o par de imagens abdômen (corte 2).	127
Figura 51 – (a) Divergência <i>Logistic loss</i> em função da divergência <i>Squared loss</i> (b) Divergência <i>Logistic loss</i> em função da divergência <i>Generalized I-divergence</i>	129
Figura 52 – Imagens S, T e diferença inicial caracterizadas pela “entropia”.	131
Figura 53 – Diferença inicial entre S e T, diferença final entre $S(x + u(x))$ e T e imagem final considerando a divergência <i>Squared loss</i> , diferença final entre $S(x+u(x))$ e T e imagem final considerando a divergência <i>Logistic loss</i> . Resultados para as caracterizações “cor”(1 ^a linha), “entropia”(2 ^a linha) and ‘cor and entropia’ (3 ^a) linha.	133
Figura 54 – Resultado caracterização “cor, entropia, desvio e média”: imagem final considerando as divergências: <i>Squared loss</i> , <i>Generalized I-divergence</i> e <i>Logistic loss</i> , respectivamente.	134

Figura 55 – (a)Evolução da variação do CC para cada par de imagem considerando as diferentes caracterizações. (b)Efeito do fator “Caracterização” sobre a variação do coeficiente de correlação. (c) Efeito do fator “Divergência” sobre a variação relativa do coeficiente de correlação.	135
Figura 56 – Caracterização “momentos”: invariante 1 e invariante 2.	136
Figura 57 – (a)Interação entre a “Caracterização” e “Divergência”. (b)Efeito do fator “Caracterização” sobre a variação do coeficiente de correlação.	137
Figura 58 – Comparação de resultados: Linha 1 - Source e Target e a diferença inicial $S(x)-T(x)$ linha 2 - resultados e diferenças finais $S(x+u)-T(x)$ obtidos pelo funcional proposto associando a caracterização “cor” com as divergências <i>Squared loss</i> e <i>Logistic loss</i> , linha 3 - resultados e diferenças finais $S(x+u)-T(x)$ obtidos pelo funcional proposto associando a caracterização “cor e entropia” com as divergências <i>Squared loss</i> e <i>Logistic loss</i> e linha 4 - resultado e diferença final obtido pela função “imregister” do Matlab	140
Figura 59 – Coeficiente de correlação x Iterações.	141
Figura 60 – Comparação de resultados: Linha 1 - Source e Target e a diferença inicial $S(x)-T(x)$ linha 2 - resultado obtido pelo método e diferença final $S(x+u)-T(x)$, linha 3 - resultado obtido pela função “imregister” do Matlab;	142
Figura 61 – Comparação de resultados: linha 1-Source e Target e a diferença inicial $S(x)-T(x)$, linha 2 - $S^{eucl}(x+u)$ e diferença final $S^{eucl}(x+u)-T(x)$, linha 3 - $S^{id}(x+u)$ e diferença final $S^{id}(x+u)-T(x)$ e linha 4 - resultado obtido pela função <i>imregister</i> do Matlab.	143
Figura 62 – Influência do fator conformal.	150
Figura 63 – Imagens do fator conformal obtidas em diferentes iterações do processo de registro.	151
Figura 64 – Imagens S, T e Diferença Inicial.	154
Figura 65 – Medidas de Avaliação do registro obtido com a divergência Total Squared loss para diferentes valores de λ	155
Figura 66 – Medidas de Avaliação do registro obtido com a divergência Squared loss para diferentes valores de lambda.	155
Figura 67 – Valores do CC em função do tempo de processamento utilizando diferentes valores para λ	156
Figura 68 – Comparação entre os métodos registro com as divergências <i>Total Squared loss</i> e <i>Squared loss</i> , utilizando λ igual a 250 e 3500, respectivamente.	157
Figura 69 – Evolução do CC para o registro com divergência <i>Total Squared loss</i> e $\lambda = 700$	158

Figura 70 – Efeito do campo de deslocamento nos pontos (73, 76), (38, 60) (70, 47) e (21, 27).	158
Figura 71 – Evolução do produto $\lambda\rho$ para os pontos: (a) (73, 76), (b) (38, 60), (c) (70, 47) e (d) (21, 27).	159
Figura 72 – Em cada linha tem-se a imagem final S^u , a diferença final $S^u - T$, o campo de deslocamento u e a superfície de deslocamento, obtidos considerando o registro com as divergências (a) <i>Total Squared loss</i> e (b) <i>Squared loss</i>	160
Figura 73 – Evolução do CC para os registros com γ fixo igual a 250 e γ variável com $\lambda = 0, 1$	161
Figura 74 – Resultados finais para os pares de imagens: próstata, coração, peito, abdômen (corte1) e abdômen (corte2) utilizando a divergência <i>Total I-divergence</i>	163
Figura 75 – Conceitos principais, contribuições, publicações e extensões.	170

Lista de tabelas

Tabela 1 – Tabela de Contigências	52
Tabela 2 – Funções Convexas unidimensionais e divergências de Bregman.	63
Tabela 3 – Divergências de Bregman Separáveis.	63
Tabela 4 – Divergências Total Bregman entre pontos unidimensionais (escalares) utilizadas para obter divergências Total Bregman Separáveis.	69
Tabela 5 – Resumo dos tratamentos definidos na seção 3.3 e notação.	72
Tabela 6 – Resultados obtidos com abordagem BoVW no banco Caltech101 com as funções Cosseno, GID/D e GID/ ε	73
Tabela 7 – Resultados obtidos no banco Caltech101 com as funções Euclidiana, Cosseno, KL e GID.	75
Tabela 8 – Funções Convexas x Equações de Euler-Lagrange.	106
Tabela 9 – Conjunto imagem de algumas divergências de Bregman em função da escolha do ϵ	112
Tabela 10 – Coeficientes de correlação obtidos para o registro das imagens S e T utilizando 200 iterações	120
Tabela 11 – Coeficientes de correlação obtidos para o registro das imagens S e T utilizando 500 iterações	121
Tabela 12 – Medidas de avaliação obtidas nos registros das imagens.	123
Tabela 13 – Tempo de processamento do método proposto utilizando 500 iterações.	124
Tabela 14 – Coeficientes de correlação final para “Caracterização” cor e “Divergen- cia” <i>Logistic loss</i> , 500 iterações, $\Delta t = 0,07$ e $\lambda = 5, 15, 30, 50$ e 60 . . .	126
Tabela 15 – Diferença entre os coeficientes de correlação inicial e final obtido con- siderando para o fator “Caracterização” a cor e para a “Divergência” a opção <i>Logistic loss</i> . 500 iterações, $\Delta t = 0,07$ e $\lambda = 5, 15, 30, 50$ e 60 . .	127
Tabela 16 – Medidas de avaliação obtidas no registro das imagens da Figura 46 com N.Iter.= 500, $\Delta = 0,07$, $\lambda = 90, 30$ e 42	128
Tabela 17 – Medidas de avaliação obtidas considerando as imagens representadas pela “entropia’- N°Iter = 1000, $\Delta t = 0,07$ e $\lambda = 50$	131

Tabela 18 – Medidas de avaliação aplicando o campo de deformação u obtido pela caracterização “entropia” às imagens iniciais S e T - $N^o Iter = 1000$, $\Delta t = 0,07$ e $\lambda = 50$	131
Tabela 19 – Comparação dos coeficientes de correlação obtidos considerando “cor” e “cor e entropia - $N^o Iter = 500$, $\Delta t = 0,07$ e $\lambda = 30$	132
Tabela 20 – Variação relativa do coeficiente de correlação - $N^o Iter = 500$, $\Delta t = 0,07$, $\lambda = 30$ e $CC_{inicial} = 0,3317$	133
Tabela 21 – Contribuição obtida em cada combinação em relação ao método tradicional: “ <i>Squared loss</i> e <i>cor</i> ” - $Iterations = 500$, $\Delta t = 0.07$, $\lambda = 30$. . .	134
Tabela 22 – Contribuição obtida pela agregação de informação em relação à característica “cor” para as divergências: <i>Squared loss</i> , <i>Logistic loss</i> , <i>Generalized I-divergence</i>	135
Tabela 23 – Variação relativa do coeficiente de correlação obtido para os registros realizados, considerando as informações cor e momentos.	136
Tabela 24 – Coeficientes de correlação dos melhores resultados obtidos utilizando as divergências <i>Logistic loss</i> e <i>Generalized I-divergence</i> associadas a outras informações além da cor.	138
Tabela 25 – Avaliação registro de diferentes imagens (sintéticas e médicas) utilizando a função “imregister”.	138
Tabela 26 – Comparação do desempenho dos métodos de registro propostos com o método disponibilizado pelo <i>software Matlab</i> - Cortes Abdominais . .	139
Tabela 27 – Comparação do desempenho do método de registro proposto com o disponibilizado pelo <i>software Matlab</i> - Cérebro.	141
Tabela 28 – Medidas de avaliação obtidas nos registros das imagens utilizando o funcional com parâmetro $\gamma = \lambda/\sigma$ e as divergências <i>Total Squared loss</i> , <i>Total Logistic loss</i> e <i>Total I-divergence</i>	162
Tabela 29 – Número de Iterações realizadas para $t=5$ min.	162
Tabela 30 – Comparação entre CC obtidos utilizando os funcionais definidos com a DB e DTB, considerando $t=5$ min e $\lambda = 0,1$	164
Tabela 31 – Comparação entre o número de iterações obtidos utilizando os funcionais definidos com a DB e DTB, considerando $t=5$ min e $\lambda = 0,1$. . .	164

Sumário

1	INTRODUÇÃO	27
1.1	Motivação	28
1.2	Trabalhos Correlatos	29
1.3	Desafios da Pesquisa e Hipóteses	31
1.4	Objetivos e Contribuições	32
1.5	Organização da Tese	33
2	FUNDAMENTAÇÃO TEÓRICA	37
2.1	Similaridade e Dissimilaridade	37
2.1.1	Dissimilaridades	38
2.1.2	Similaridade	39
2.2	Agrupamentos por similaridade	40
2.2.1	Métodos Hierárquicos Aglomerativos	40
2.2.2	Métodos de particionamento iterativo	41
2.2.3	Validação dos Agrupamentos	42
2.3	Segmentação hierárquica	45
2.3.1	Medidas de Avaliação	48
2.4	Recuperação de imagens baseada em conteúdo	49
2.4.1	Medidas de avaliação	51
2.4.2	Precisão em k	53
2.4.3	<i>Mean Average Precision</i>	53
2.5	Registro variacional de imagens	53
2.5.1	Equações de Euler-Lagrange	55
2.5.2	Método das Diferenças Finitas	57
2.5.3	Interpolação	58
2.5.4	Medidas de avaliação	58

3	DIVERGÊNCIAS DE BREGMAN E TOTAL BREGMAN:	
	DEFINIÇÃO, DOMÍNIO E TRATAMENTOS	61
3.1	Divergência de Bregman	62
3.1.1	Informação de Bregman	64
3.1.2	Perda da Informação de Bregman	64
3.1.3	Árvores de Bolas de Bregman (<i>BB-tree</i>)	65
3.2	Divergência Total Bregman	67
3.2.1	t-Centro	69
3.3	Tratamento dos dados para a aplicação das Divergências de Bregman e Total Bregman	70
3.3.1	Avaliação dos tratamentos apresentados	72
3.4	Considerações Finais	77
4	ÁRVORE TOTAL BREGMAN	79
4.1	Introdução	79
4.2	Bolas Total Bregman	81
4.2.1	Projeção na bola Total Bregman	82
4.3	TBB-<i>tree</i>	83
4.3.1	Busca com TBB- <i>tree</i>	84
4.4	Aplicações	85
4.4.1	<i>Clustering</i>	85
4.4.2	Re-ranking	88
4.5	Considerações finais	90
5	DIVERGÊNCIA DE BREGMAN APLICADA EM PROBLEMAS DE SEGMENTAÇÃO HIERÁRQUICA	91
5.1	Método Proposto	92
5.1.1	Custo de Fusão	93
5.1.2	Critério de Fusão	96
5.2	Detalhes da implementação	96
5.3	Experimentos e Análise dos Resultados	97
5.4	Considerações Finais	98
6	DIVERGÊNCIAS DE BREGMAN APLICADAS NO REGISTRO NÃO RÍGIDO DE IMAGENS	101
6.1	Formulação do modelo de registro	103
6.1.1	Fluxo do gradiente das divergências de Bregman	105
6.1.2	Caracterização utilizando informações locais	106
6.2	Detalhes da implementação	107
6.3	Experimentos e Análise de Resultados	108

6.3.1	Domínio de aplicação das divergências de Bregman	112
6.3.2	Avaliação da convergência do método proposto para diferentes opções de “Divergência” e “Caracterização”	112
6.3.3	Análise da influência do fator “Divergência” no registro das imagens.	122
6.3.4	Análise do impacto do parâmetro lambda.	126
6.3.5	Análise do efeito causado pela agregação de características	130
6.3.6	Comparação de métodos	137
6.4	Considerações finais	144
7	DIVERGÊNCIAS TOTAL BREGMAN APLICADAS NO REGISTRO NÃO RÍGIDO DE IMAGENS	147
7.1	Modelo de registro com Divergências Total Bregman	148
7.1.1	Detalhes da implementação	149
7.2	Fator conformal	150
7.3	Modelo de registro com Divergências Total Bregman e parâmetro de balanceamento auto-ajustável	151
7.4	Detalhes da implementação	153
7.5	Experimentos e Análise de Resultados	154
7.5.1	Experimento I	154
7.5.2	Experimento II	158
7.5.3	Experimento III	159
7.6	Considerações finais	164
8	CONCLUSÃO	167
8.1	Produtos gerados	168
8.2	Trabalhos futuros	169
REFERÊNCIAS		171

CAPÍTULO 1

Introdução

A análise de imagens digitais e as técnicas utilizadas na visão computacional envolvem a manipulação de dados complexos que não são representados por um escalar, mas sim por elementos de um espaço definido por suas características. Nesta perspectiva, nas aplicações que envolvem recuperação, classificação, agrupamento, segmentação e registro faz-se necessário o estudo da similaridade ou dissimilaridade entre dados que são representados por vetores, matrizes ou funções de probabilidade, dentre outras.

A similaridade é um valor numérico que representa o quanto dois objetos ou dados são semelhantes, isto é, quanto maior for o valor observado, mais parecidos serão os objetos. Enquanto que a dissimilaridade associa a cada par de dados um número que pode ser interpretado como a discrepância entre eles. Desta forma, quanto maior for a medida de dissimilaridade menor será a semelhança entre os indivíduos.

A dissimilaridade pode ser definida por diferentes funções de divergências. As divergências podem não ser, necessariamente, uma métrica (distância). As distâncias Euclidiana e de Mahalanobis são exemplos de divergências que atendem às propriedades de uma métrica. Entretanto, uma medida de dissimilaridade pode ser bastante eficaz sem ser uma métrica. Por exemplo, divergência Kullback-Leibler (KL) usada para comparar duas funções de densidade de probabilidade, (XU et al., 2012), não é simétrica e não obedece à desigualdade triangular (ZEZULA et al., 2006).

As distâncias Euclidiana (MONTGOMERY; PECK; VINING, 2012) e de Mahalanobis (MAESSCHALCK; JOUAN-RIMBAUD; MASSART, 2000), e a divergência Kullback-Leibler(KL) são casos especiais da Divergência de Bregman (DB), introduzida por Bregman em 1967 (BREGMAN, 1967). Essas divergências são medidas de dissimilaridade obtidas a partir de funções convexas e apresentam propriedades importantes para as aplicações na área de processamento de imagens.

Este trabalho tem como foco o estudo das divergências de Bregman e das divergências Total Bregman(DTB), definidas em (LIU, 2011) a partir do conceito das divergências de Bregman. ’

1.1 Motivação

A maior parte dos trabalhos que são dedicados ao desenvolvimento de estruturas e algoritmos que aplicam medidas de dissimilaridade, utilizam espaços métricos, como por exemplo, o Espaço Euclidiano (ZHANG et al., 2009; DATTA et al., 2008; GAO et al., 2015). Nos espaços métricos as funções de dissimilaridade são estritamente simétricas, não negativas e satisfazem a desigualdade triangular. Entretanto, existem aplicações reais que envolvem a comparação de probabilidades, matrizes, séries temporais, imagens, dentre outras, nas quais a noção de dissimilaridade adequada não pode ser definida por uma métrica.

Por exemplo, nos problemas de recuperação de imagens, ao realizar uma busca na qual o objeto de consulta é parte de uma imagem, verifica-se uma dissimilaridade entre a parte e o todo, pois a parte está contida no todo. Entretanto, se a imagem toda for o objeto de consulta, a dissimilaridade estabelecida entre a imagem toda e sua parte deve ser alta. As funções convencionais atribuem a mesma medida de dissimilaridade entre “toda imagem” e “parte da imagem” nas duas situações.

Assim, embora os métodos desenvolvidos para a análise de imagens e visão computacional tenham gerado bons resultados, existem situações para as quais as funções de dissimilaridades convencionais apresentam limitações. Devido às suas propriedades, as DBs e DTBs surgem como alternativas para medir a dissimilaridade e proximidade nestas aplicações.

As DBs e DTBs se diferenciam das métricas convencionais pois não satisfazem sempre a desigualdade triangular e podem não ser simétricas. Além disso, possuem propriedades importantes na busca por similaridades que são válidas para uma família de funções, por exemplo, o argumento mínimo da divergência de Bregman para um conjunto de dados n-dimensional é a média desse conjunto, independentemente da escolha da divergência (BANERJEE et al., 2005a) e o centro obtido com qualquer divergência total de Bregman é robusto a dados ruidosos e atípicos (LIU, 2011). A DTB é ainda, invariante em relação a movimentos rígidos (translação e rotação).

Alguns algoritmos e estruturas importantes foram generalizados utilizando as DBs (BANERJEE et al., 2005a; DASGUPTA; TELGARSKY, 2012; CAYTON, 2008). A generalização possibilita a escolha da medida de dissimilaridade tornando flexível a aplicação dos métodos que envolvem a busca por similaridade. Em muitas aplicações não está claro qual medida é mais adequada e em diversos trabalhos a escolha da medida é posta como um problema de aprendizagem (XING et al., 2002; GOLDBERGER et al., 2004; WEINBERGER; SAUL, 2009).

A generalização propiciada pela definição das divergências de Bregman e Total Bregman possibilita o desenvolvimento de métodos que sejam eficientes e aplicáveis considerando um conjunto de funções de dissimilaridades ao invés de uma única medida. A vantagem desta abordagem é a possibilidade de escolher uma medida de dissimilaridade

adequada ao problema em questão.

Apesar da flexibilidade proporcionada pelas divergências a sua aplicação têm alguns desafios. Muitos dos métodos já desenvolvidos utilizam em seus procedimentos propriedades geométricas e algébricas de funções definidas como métricas o que torna inviável a aplicação direta das DB e DTB. Os algoritmos que, por exemplo, são construídos considerando as propriedades da norma l_2 , podem não ser diretamente aplicáveis utilizando outras funções de dissimilaridade tal como o algoritmo de classificação *k-Nearest Neighbor* (*k* vizinhos mais próximos) que é baseado, principalmente, na propriedade da desigualdade triangular que pode não ser satisfeita pela divergências de Bregman, (JAIN; MURTY; FLYNN, 1999).

Além disso, algumas divergências de Bregman e Total Bregman são definidas utilizando a função logarítmica cujo domínio é o conjunto de valores reais positivos, por outro lado a caracterização das imagens, geralmente, é dada por vetores que podem assumir valores iguais a zero em suas coordenadas. Desta forma, a aplicação das divergências exige um tratamento dos valores desses vetores de representação a fim de assegurar o domínio dessas funções.

Sendo assim, para avaliar a dissimilaridade dos dados utilizando as divergências de Bregman e Total Bregman é preciso criar condições que possibilitem aplicá-las adequadamente considerando as propriedades destas divergências e o seu domínio de definição.

Neste trabalho propomos tratamentos e novas formulações que possibilitem a aplicação de alguns métodos de análise de imagens considerando as divergências de Bregman e Total Bregman como medida de dissimilaridade.

1.2 Trabalhos Correlatos

Diversas pesquisas têm sido realizadas no intuito de desenvolver métodos eficientes para uma família de funções de distância que partilham de propriedades semelhantes. Algumas destas pesquisas têm apresentado resultados e generalizações substituindo a distância Euclidiana pelas divergências de Bregman na resolução de problemas de otimização. Dentro destes, pode-se destacar a generalização do método de ponto proximal apresentada em (KIWIEL, 1997). Em (BECK; TEBOULLE, 2003) e (CHEN; TEBOULLE, 1993), as DBs foram exploradas em problemas de estrutura. Destaca-se, ainda, a aplicação em problemas de agrupamento em tomografia computadorizada, (BEN-TAL; MARGALIT; NEMIROVSKI, 2001) e distribuições de famílias exponenciais, (BANERJEE et al., 2005b).

Recentemente as DBs têm contribuído para avanços nas pesquisas da área de processamento de imagens, envolvendo aplicações de agrupamentos (BANERJEE et al., 2005b), (DASGUPTA; TELGARSKY, 2012), segmentação (GOLDSTEIN; BRESSON; OSHER, 2010), classificação (XU et al., 2012). Em (GOLDSTEIN; BRESSON; OSHER, 2010), os

autores obtêm a segmentação de imagens aplicando o método *Split Bregman* desenvolvido para problemas de regularização L_1 .

Algoritmos fundamentais como *k-means* (JAIN; MURTY; FLYNN, 1999), PCA (BUNTINE, 2002) e estruturas de dados tais como: os diagramas de Voronoi (SCHWANDER; NIELSEN, 2010), foram generalizados para a classe das divergências de Bregman. Por exemplo, em Banerjee et al. os autores mostraram que o algoritmo *k-means*, de Lloyd, pode ser estendido para classe de divergências Bregman unificando vários algoritmos. Neste contexto, em (DASGUPTA; TELGARSKY, 2012), Dasgupta e Telgarky provam utilizando o conceito de divergência de Bregman, que o custo de fusão “*k-means*” é uma consequência natural do algoritmo de agrupamentos *hard* apresentado em (BANERJEE et al., 2005b) e desenvolvem uma teoria de agrupamentos aglomerativa com divergências de Bregman. O custo de fusão é então generalizado permitindo a escolha de diferentes divergências de Bregman.

Lawrence Cayton apresentou em (CAYTON, 2008) propriedades geométricas e algébricas necessárias para a extensão dos métodos de busca por vizinhos próximos mediante a utilização das divergências de Bregman. Este trabalho apresenta um algoritmo de busca de vizinhos mais próximos baseado na decomposição hierárquica do espaço e em corpos convexos denominados bolas de Bregman (BB). A estrutura de dados obtida é denominada árvore de bolas de Bregman (*bb-tree*).

O método apresentado em (XU et al., 2012) aplica as DBs em problemas de recuperação de imagens baseada em conteúdo (*Content-based image retrieval – CBIR*). Os autores propõem uma nova formulação para “*manifold ranking*” e apresentam duas extensões que melhoraram a performance do ranqueamento. As principais contribuições são: Proposta de um algoritmo de ranqueamento de “*manifolds*” como um problema de otimização utilizando divergência de Bregman; Apresentação de um novo entendimento no contexto de MRs (*manifold ranking*) introduzindo “aprendizado” utilizando divergência de Bregman e duas extensões eficientes e efetivas que melhoraram a performance da tradicional MR.

Em (LIU et al., 2012), uma nova classe de divergência denominada divergência Total Bregman (DTB), é definida. Na literatura encontram-se aplicações das DTB na recuperação da forma (LIU et al., 2011), na análise de imagens de Tensor de Difusão (DTI) (LIU, 2011), em métodos de aprendizagem de máquina (LIU et al., 2012) e em técnicas de rastreamento de objetos (ROMERO; LACASSAGNE; GOUIFFES, 2013).

Nas pesquisas que envolvem registro de imagens, alguns trabalhos, tais como (LI et al., 2014), (MARTIN; DURRANI, 2007), (HE; HAMZA; KRIM, 2003) e (GUETTER et al., 2005) utilizam diferentes medidas de divergências. Em (ALVAREZ et al., 2007) têm-se modelos variacionais com diferentes medidas de dissimilaridade. Entretanto, as divergências de Bregman e Total Bregman não foram utilizadas na generalização dos métodos.

Embora as DBs e DTBs tenham sido exploradas em diversas aplicações, ainda é restrita

a sua utilização na generalização de métodos utilizados na análise de imagens. Neste contexto, destacam-se as generalizações de métodos de agrupamentos com divergências de Bregman e das árvores de Bolas.

Considerando a relevância das generalizações de métodos utilizando as DBs e DTBs, neste trabalho é proposta a generalização do método de segmentação baseado no funcional de Mumford Shah utilizando as DBs e da construção de árvores de bolas utilizando as divergência Total de Bregman.

Propõe-se ainda a generalização dos métodos variacionais aplicados no registro de imagens. Os trabalhos existentes aplicando as DBs e DTBs no contexto de imagens diferem ainda, da proposta deste projeto, no que se refere à aplicação destas divergências assegurando, por meio de tratamentos adequados, o seu domínio de aplicação.

1.3 Desafios da Pesquisa e Hipóteses

As divergências de Bregman e Total Bregman são úteis para determinar a dissimilaridade de dados complexos e têm sido utilizadas em diversas aplicações (BANNERJEE et al., 2005a; LIU et al., 2012; DASGUPTA; TELGARSKY, 2012; CAYTON, 2008). Algoritmos fundamentais e estruturas de dados foram generalizados oferecendo, assim, meta-algoritmos que podem ser aplicados utilizando qualquer divergência de Bregman. Lawrence Cayton apresenta um algoritmo de busca por vizinhos mais próximos que não utiliza a desigualdade triangular e baseia-se na decomposição hierárquica do espaço em corpos convexos denominados bolas de Bregman (BB). A estrutura de dados obtida é denominada *árvore de bolas de Bregman (BB-tree)* (CAYTON, 2009). Não existe na literatura, uma estrutura que divida o espaço de busca em bolas definidas pela divergência Total Bregman, devido às propriedades desta divergência a análise e avaliação dessa possibilidade é considerada neste trabalho.

Nesta perspectiva surgem as seguintes questões: “É possível que os métodos utilizados no processamento de imagens se tornem mais eficazes ao utilizar como medida de dissimilaridade as DB e DTB?”, “Quais tratamentos devem ser utilizados para que estas medidas possam ser aplicadas adequadamente?”, “Seria possível que outros métodos e estruturas utilizados no processamento e análise de imagens sejam generalizados possibilitando a sua aplicação com diversas divergências? e “Em quais situações, as DB e DTB fornecem melhores resultados quando comparados àqueles obtidos com funções métricas?”.

Desta forma, tendo em vista as questões mencionadas, consideraram-se as seguintes hipóteses:

- Os métodos de segmentação hierárquica que utilizam o funcional de Mumford Shah e os métodos variacionais aplicados no registro de imagens podem ser generalizados utilizando as divergências de Bregman e Total Bregman;

- As propriedades geométricas e algébricas das divergências Total Bregman possibilitam a criação de uma nova estrutura de dados, semelhante à árvore de bolas de Bregman, eficaz na busca por vizinhos próximos;
- A definição de tratamentos que considerem as características dos dados e os domínios de aplicação das divergência propicia um melhor eficácia dos métodos que utilizam as divergências de Bregman e Total Bregman.

1.4 Objetivos e Contribuições

Este trabalho tem como foco principal propor alternativas que propiciem o uso abrangente e eficaz das divergências de Bregman e Total Bregman como funções de dissimilaridade, utilizadas na análise de imagens digitais e na visão computacional. Os objetivos principais são:

- Propor uma generalização do método de segmentação hierárquica baseado no funcional de Mumford Shah utilizando o conceito de Perda de Informação de Bregman e a teoria de Clusterização aglomerativa com divergências de Bregman;
- Apresentar e investigar, considerando as características dos dados e os domínios de aplicação das divergências, possíveis tratamentos que possibilitem a utilização das divergências de Bregman e Total Bregman em aplicações empregadas na análise de imagens;
- Combinar a estrutura das BB-trees com o conceito de divergência Total Bregman, analisando a sua aplicação no contexto da recuperação de imagens por conteúdo e agrupamentos;
- Investigar o uso das divergências de Bregman e Total Bregman em registro de imagens.

As principais contribuições deste trabalho são:

- Um método de segmentação hierárquica que é uma generalização do método de Mumford-Shah, possibilitando a segmentação para diferentes divergências de Bregman;
- Tratamentos adequados à utilização das divergências de Bregman e Total de Bregman como medidas de dissimilaridade no contexto da recuperação de imagens e agrupamentos;
- Uma estrutura de dados utilizando vizinhanças definidas com a divergência Total de Bregman que seja eficaz na recuperação de imagens e agrupamentos.

- Funcionais que possibilitem a aplicação das divergências de Bregman e Total de Bregman em problemas de registro de imagens.

1.5 Organização da Tese

Esta tese é organizada de acordo com as principais contribuições obtidas em nossa pesquisa, as quais foram publicadas ou submetidas a conferências e revistas. A seguir, descrevemos o conteúdo de cada capítulo indicando a publicação ou submissão:

Capítulo 2: Este capítulo apresenta conceitos que alicerçam a utilização das divergências de Bregman e Total Bregman como medidas de dissimilaridade no contexto da recuperação, classificação, agrupamento, segmentação e registro de imagens;

Capítulo 3: Neste capítulo são apresentadas definições e propriedades das divergências. Ainda, considerando o domínio destas divergências são definidos e avaliados tratamentos que possibilitam a utilização destas no contexto da análise de imagens. Parte desse capítulo pode ser encontrada no artigo publicado em *Electronics, Circuits and Systems (ICECS), 2014 21st IEEE International Conference on* (pp. 734-737). *IEEE*. (ROCHA; FERREIRA; BARCELOS, 2014);

Capítulo 4: Este capítulo apresenta um novo funcional baseado no conceito da Informação de Bregman o qual é utilizado para a construção de um método de segmentação hierárquica. Este conteúdo pode ser encontrado no artigo publicado em *Ibero-american Congress on Pattern Recognition (pp. 493-500). Springer International Publishing*. (FERREIRA; BACKES; BARCELOS, 2015);

Capítulo 5: Neste capítulo é apresentado uma nova estrutura de dados e um novo algoritmo para *Nearest neighbor search* utilizando a divergência Total Bregman. O conteúdo do capítulo foi submetido na forma de artigo à revista *Neurocomputing* e em parte pode ser encontrado no artigo publicado em *Proceedings of the Symposium on Applied Computing (pp. 120-124). ACM*. (FERREIRA; ROCHA; BARCELOS, 2017);

Capítulo 6: Este capítulo apresenta um novo funcional de energia que representa o problema de registro de imagens utilizando as divergências de Bregman. O conteúdo deste capítulo foi submetido à revista *Pattern Recognition* e parte pode ser encontrado no artigo publicado em *Proceedings of the Symposium on Applied Computing (pp. 158-163). ACM*. (FERREIRA; RIBEIRO; BARCELOS, 2017);

Capítulo 7: Este capítulo é dedicado à aplicação das divergências total Bregman ao método de registro variacional. São propostos dois novos funcionais, no primeiro

funcional é utilizada como termo de similaridade a imagem residual dada pela divergência Total Bregman entre as imagens envolvidas no registro. Em nossa segunda proposta, o novo funcional representa o problema de registro de imagens a partir das divergências Total Bregman utilizando como termo de similaridade a função log-verossimilhança negativa da imagem residual;

Conclusão: Este capítulo apresenta um resumo comentado dos principais resultados e discute a viabilidade da aplicação dos métodos propostos.

A organização geral desta tese considerando os principais conceitos, contribuições e publicações é ilustrada na Figura 1.

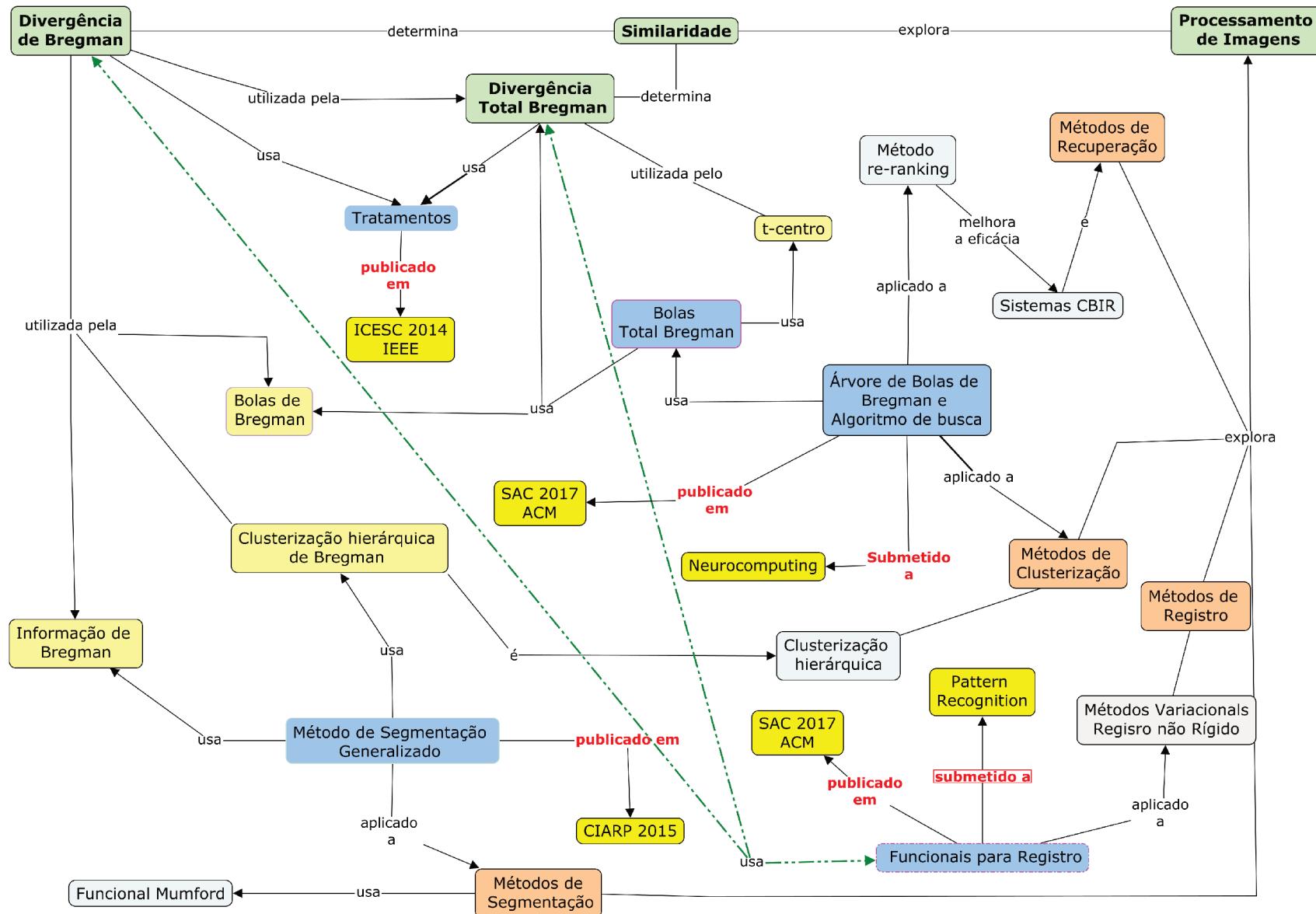


Figura 1 – Conceitos principais, contribuições e publicações.

CAPÍTULO 2

Fundamentação Teórica

Tendo como objetivo a utilização das divergências de Bregman e Total Bregman como medidas de dissimilaridade no contexto da recuperação, classificação, agrupamentos, segmentação e registro de imagens, apresentam-se neste capítulo os seguintes conceitos: Similaridade e dissimilaridade, Agrupamentos por similaridade, Segmentação hierárquica, Recuperação de imagens baseada em conteúdo e Registro de imagens.

2.1 Similaridade e Dissimilaridade

Diversas técnicas utilizadas na análise de imagens envolvem o estudo da proximidade entre objetos. A proximidade entre dois objetos é definida em função da similaridade ou dissimilaridade entre os atributos correspondentes destes objetos.

Informalmente, a similaridade entre dois objetos é uma medida numérica que representa o quanto dois objetos são iguais. Desta forma, a similaridade será maior para pares de objetos que são mais parecidos. A similaridade é não-negativa e frequentemente entre 0 (sem similaridade) e 1 (semelhança completa).

A dissimilaridade entre dois objetos é uma medida numérica do grau de diferença entre esses objetos. As diferenças são menores para pares de objetos mais semelhantes. Frequentemente, o termo distância é usado como sinônimo de dissimilaridade, embora, como veremos, as distâncias são uma classe especial de medidas de dissimilaridades. As dissimilaridades podem pertencer ao intervalo $[0,1]$, mas também é comum que elas variem de 0 a ∞ .

As medidas de proximidade, especialmente as de similaridade, podem ser definidas de modo que assumam valores no intervalo $[0, 1]$. Assim, o valor de proximidade indicará a fração de similaridade (ou dissimilaridade) entre dois objetos. De um modo geral, considerando s o valor original da similaridade e s' o novo valor, pode-se utilizar a transformação: $s' = (s - \min_s) / (\max_s - \min_s)$, onde \max_s e \min_s são os valores de máximo e mínimo para a similaridade, respectivamente. De modo análogo as funções de dissimilaridade

podem ser mapeadas para o intervalo $[0, 1]$ usando $d' = (d - \min_d) / (\max_d - \min_d)$ sendo \max_d e \min_d os valores de máximo e mínimo para a dissimilaridade, respectivamente.

Entretanto, pode haver várias complicações no mapeamento de medidas de proximidade no intervalo $[0, 1]$. Se, por exemplo, a medida de proximidade toma valores originalmente no intervalo $[0, \infty]$, será necessária uma transformação não linear. Pode-se considerar a transformação $d' = d / (1 + d)$ para uma medida de dissimilaridade que varia de 0 a ∞ . Observe que, os valores não terão o mesmo relacionamento entre si nesta nova escala. Valores maiores na escala de dissimilaridade original são compactados na faixa de valores próximos a 1. Outra complicação é que o significado da medida de proximidade pode ser alterada. Por exemplo, a medida de similaridade correlação recebe valores no intervalo $[-1, 1]$. Mapeando estes valores no intervalo $[0, 1]$ perde-se informação sobre o sinal que pode ser importante em algumas aplicações. Desta forma, deve-se avaliar o impacto dessas consequências no estudo em questão (TAN et al., 2006).

Um valor de similaridade pode ser facilmente transformado em dissimilaridade e vice-versa, embora deva-se, também, considerar as questões da preservação de significado e da mudança de uma escala linear para uma escala não linear. Se a similaridade (ou dissimilaridade) pertencer ao intervalo $[0, 1]$, então a dissimilaridade (ou similaridade) pode ser definida como $d = 1 - s$ ($s = 1 - d$). Outra abordagem simples é definir a similaridade como o negativo da dissimilaridade (ou vice-versa). Para ilustrar, as dissimilaridades 0, 1, 10 e 100 podem ser transformadas nas similaridades 0, -1, -10 e -100, respectivamente. Em geral, qualquer função monotônica decrescente pode ser usada para converter dissimilaridades em similaridades, ou vice-versa (TAN et al., 2006).

2.1.1 Dissimilaridades

A dissimilaridade pode ser definida por diferentes funções de divergências. Essas funções podem ser métricas (distância) ou não. Sendo X um conjunto qualquer, uma métrica é uma função $d : X \times X \rightarrow \mathbb{R}$ que satisfaz as propriedades:

1. Positividade: $d(x, y) \geq 0 \quad \forall x, y \in X$
2. Condição de distância nula: $d(x, y) = 0 \iff x = y$
3. Simetria: $d(x, y) = d(y, x) \quad \forall x, y \in X$
4. Desigualdade Triangular: $d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z \in X,$

Neste caso, $d(x, y)$ é denominada distância entre x e y e o par (X, d) é chamado espaço métrico (ZEZULA et al., 2006).

A proximidade entre atributos contínuos é mais frequentemente expressa em termos de dissimilaridades, e as distâncias fornecem uma maneira bem definida de combinar essas diferenças em uma medida geral de proximidade.

As distâncias Euclíadiana, *Manhattan* e *City Block* são exemplos de métricas. No entanto, muitas dissimilaridades não satisfazem uma ou mais das propriedades métricas tais como as divergências Kullback-Leibler (KL), Itakura Saito e logistic loss.

As divergências Bregman e total Bregman, objeto de estudo deste trabalho, são famílias de funções de proximidade que compartilham propriedades comuns. Estas divergências são funções de perda ou distorção definidas . Quanto mais próximos forem os dados, menor a perda ou distorção. Desta forma as divergências Bregman e total Bregman são famílias de funções de dissimilaridade. As divergências de Bregman e total Bregman envolvem tanto funções métricas quanto não métricas.

2.1.2 Similaridade

Considerando que $s(x, y)$ é a similaridade entre os dados x e y , então s satisfaz as propriedades:

1. $s(x, y) = 1 \iff x = y \quad (0 \leq s \leq 1).$
2. $s(x, y) = s(y, x), \forall x \text{ e } y$ (Simetria).

Não existe uma propriedade análoga à desigualdade triangular para medidas de similaridade. Entretanto às vezes é possível, converter uma medida de similaridade em uma distância métrica, tais como as medidas de similaridade cosseno e Jaccard. Além dessas, são exemplos de medida de similaridade: o coeficiente de correlação de Pearson, o coeficiente *Simple Matching* (SMC), o coeficiente Tanimoto (TAN et al., 2006).

Para dados esparsos, que geralmente consistem em atributos assimétricos, é comum utilizar medidas de similaridade que ignoram correspondências entre dois valores iguais a zero. Conceitualmente, isso reflete o fato de que, para um par de objetos complexos, a similaridade depende do número de características que ambos compartilham, em vez do número de características que os diferenciam. Mais especificamente, para dados esparsos e assimétricos, a maioria dos objetos tem apenas algumas das características descritas pelos atributos e, portanto, são altamente semelhantes em termos das características que não possuem. As medidas cosseno, Jaccard e Jaccard estendida são apropriadas para tais dados (TAN et al., 2006).

A escolha apropriada de uma medida de proximidade pode ser uma tarefa demorada que requer uma consideração cuidadosa do conhecimento do domínio e do propósito para o qual a medida está sendo usada. Uma série de diferentes medidas de similaridade pode precisar ser avaliada para ver quais produzem resultados que fazem mais sentido para dados específicos.

2.2 Agrupamentos por similaridade

Os algoritmos de agrupamento (*clustering*) visam organizar os elementos em grupos (*clusters*) de elementos semelhantes. Os grupos formados devem apresentar alto grau de similaridade entre seus elementos internos e baixo grau de similaridade em relação aos elementos pertencentes a outros grupos, (JAIN; MURTY; FLYNN, 1999).

Quanto à sua disposição, os grupos podem ser hierárquicos ou não hierárquicos. Segundo Hartigan, os grupos hierárquicos possuem uma estrutura de árvore, estabelecendo portanto, uma relação de hierarquia entre os grupos formados. Já os agrupamentos não-hierárquicos são agrupamentos constituídos por partições isoladas ou disjuntas, sem nenhuma relação de hierarquia. Sendo assim os processos de agrupamento podem ser divididos em dois grupos: métodos hierárquicos (*hierarchical clustering*) e métodos de particionamento (*partitioning clustering*). A seguir apresenta-se uma descrição dos métodos hierárquicos obtidos por aglomeração e do métodos de particionamento iterativos.

2.2.1 Métodos Hierárquicos Aglomerativos

Os métodos hierárquicos aglomerativos possuem uma abordagem “*bottom-up*”, ou seja, iniciam seu processamento considerando que cada elemento corresponde a um grupo e então unem pares de grupos sucessivamente até que todos os elementos tenham sido unidos em um único grupo.

Como resultado, esses métodos produzem uma estrutura em árvore binária, denominada dendrograma. Um exemplo de um dendrograma pode ser visto na Figura 2 na qual o eixo vertical representa o valor de similaridade com o qual os grupos foram unidos. Os nós folhas do dendrograma correspondem aos elementos do conjunto de dados que está sendo agrupado. Os algoritmos hierárquicos aglomerativos utilizam uma matriz de similarida-

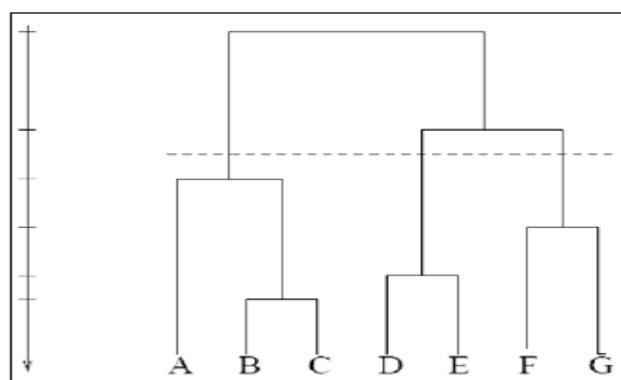


Figura 2 – Exemplo de Dendograma.
Fonte: (JAIN; MURTY; FLYNN, 1999).

des para realizar o agrupamento entre dois elementos. Essa matriz é obtida por meio do

cálculo do valor de similaridade entre todos os elementos do conjunto de dados sobre o qual será realizado o processo de agrupamento.

Os algoritmos de agrupamentos hierárquicos se diferenciam pela forma com que a similaridade entre dois grupos é calculada, a seguir apresentam-se alguns destes métodos, (ALDENDERFER; BLASHFIELD, 1984):

- Ligação simples (*single linkage*): esse método calcula a similaridade entre dois grupos como sendo o maior valor de similaridade entre os elementos pertencentes a esses grupos. Esse método tem uma tendência de encadear elementos criando grupos alongados o que o torna inapropriado para isolar grupos fraca mente separados.
- Ligação completa (*complete linkage*): diferentemente do método de ligação simples, esse método calcula a similaridade entre dois grupos como sendo o menor valor de similaridade entre os elementos pertencentes a esses grupos. Dessa forma, diz-se que esse método utiliza a regra do vizinho mais distante. Esse método tem a tendência de gerar agrupamentos menores, compactos e hiper-esféricos, contendo elementos altamente similares.
- Ligação pelo valor médio ou ligação mediana (*average linkage*): esse método calcula a similaridade como sendo a média da similaridades entre o elemento a ser adicionado ao grupo e os demais elementos já presentes no grupo. O novo elemento é adicionado ao grupo cuja similaridade média for maior. Essa similaridade média pode ser calculada de várias formas, sendo que a forma mais utilizada é a média aritmética.
- Método de *Ward*: esse método é uma variação dos anteriores e procura unir, a cada iteração, o par de grupos cuja união otimiza a variância entre os grupos, juntando os elementos cuja soma dos quadrados entre eles seja mínima ou que o erro desta soma (denominada ESS - *Error Sum of Squares*) seja mínimo. Tende a produzir grupos hiperesféricos de tamanhos muito semelhantes.

2.2.2 Métodos de particionamento iterativo

Os métodos de particionamento buscam encontrar, iterativamente, a melhor partição dos n objetos em k grupos. Frequentemente, os k grupos encontrados pelos métodos de particionamento são de melhor qualidade (grupos internamente mais homogêneos) do que os k produzidos pelos métodos hierárquicos, (NG; HAN, 1994).

O algoritmo mais conhecido dessa categoria é o *k-means*, k corresponde ao número de grupos a ser obtido pelo algoritmo. A cada iteração, os objetos são agrupados em função do centro do grupo mais próximo e, por consequência, os centros dos grupos são reavaliados. Isto provoca no espaço de atributo um deslocamento dos centros médios. O algoritmo é interrompido quando as médias não mais são deslocadas, ou há uma insignificante relocação de objetos entre os grupos (ZHANG; HSU; DAYAL, 2001).

Outro grupo de métodos de particionamento são os *k-medoids*. A diferença básica em relação ao *k-means* está na utilização de um objeto representativo, chamado de “*medoid*”, localizado mais ao centro possível do grupo, ao invés de um centro médio (KAUFMAN; ROUSSEEUW, 1990).

2.2.3 Validação dos Agrupamentos

A avaliação dos resultados de agrupamento, com o intuito de determinar o que melhor se ajusta aos dados, consiste em uma das questões mais importantes e difíceis no agrupamento de dados.

A validação do agrupamento inclui medidas baseadas em critérios externos, internos e relativos. As medidas de validação baseadas em critérios externos avaliam o grau de correspondência entre o agrupamento obtido e uma partição de referência, dos mesmos dados, conhecida ou esperada. Enquanto que os índices baseados em critérios internos avaliam o grau de compatibilidade dos grupos considerando apenas os dados e os agrupamentos obtidos. A validação utilizando critérios relativos analisa os agrupamentos com respeito a algum aspecto como por exemplo estabilidade. Geralmente é utilizado determinar os parâmetros dos algoritmos (ex. número de grupos), (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2001).

Validação baseada em dados externos

Nessa categoria encontram-se as seguintes medidas: entropia, pureza, precisão, revo- cação, medida F (*F-measure*) ou média harmônica, índice de Jaccard, estatística de Rand (*Rand statistic*), índice de Folkes e Mallows, estatística de Huberts (G) e estatística de Huberts normalizada (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2001). A partir dessas medidas, pode-se determinar se o resultado do processo de agrupamento por similaridade aproxima-se mais da estrutura pré-definida, e, portanto, estabelecer se determinado algoritmo é mais eficiente do que outro, (TAN; STEINBACH; KUMAR, 2005).

Sejam $C = \{c_1, c_2, \dots, c_n\}$ a partição de referência de um conjunto de dados X e $P = \{p_1, p_2, \dots, p_k\}$ o agrupamento a ser avaliado. Por conveniência os agrupamentos c_i da partição de referência serão denominados *classes* e os grupos p_i definidos na partição P , grupos. Os critérios apresentados a seguir são baseados na comparação de pares de pontos, (x_i, x_j) , em que x_i e $x_j \in X$. Considere,

- ❑ a = Número de pares formados por pontos que pertencem à mesma classe e ao mesmo grupo;
- ❑ b = Número de pares cujos pontos pertencem à mesma classe e a grupos distintos;
- ❑ c = Número de pares cujos pontos pertencem à classes distintas e ao mesmo grupo;
- ❑ d = Número de pares com pontos pertencentes a classes distintas e a grupo distintos.

Observe que, se M é o total de pares possíveis obtidos com os pontos do conjunto de dados X , $a + b + c + d = M$. Assim, definem-se os seguintes índices de similaridade entre os agrupamentos C e P . A Figura 3 apresenta um exemplo do cálculo das grandezas a, b, c e d .

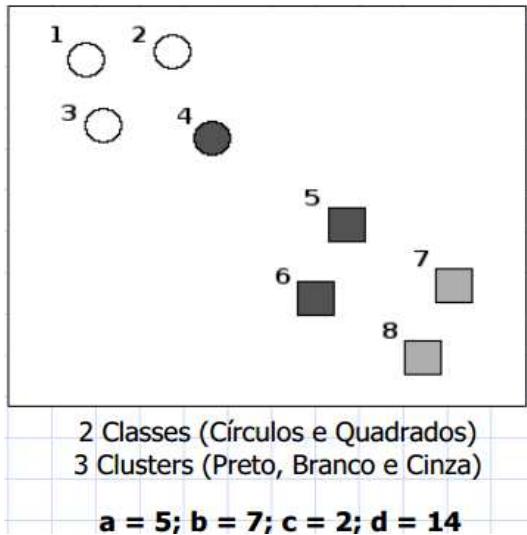


Figura 3 – Exemplo do cálculo de a, b, c e d .
Fonte: (VENDRAMIN; CAMPELLO; HRUSCHKA, 2010).

Algumas medidas de validação externa são definidas a seguir:

- ❑ Estatística de Rand - $R = \frac{a+b}{a+b+c+d}$;
- ❑ Índice de Jaccard - $J = \frac{a}{a+b+c}$ e
- ❑ Índice de Folkes e Mallows - $FM = \sqrt{\frac{a}{a+b} \cdot \frac{a}{a+c}}$.

Quanto maior os índices R , J e FM maior é o grau de correspondência entre o agrupamento obtido P e uma partição de referência, C .

Validação baseada em dados internos

Os métodos de validação baseados em critérios internos utilizam apenas as informações contidas nos grupos gerados para realizar a validação dos resultados, isto é, não utilizam informações externas, como a estrutura de classes utilizada pelos métodos baseados em critérios externos. O cálculo destes índices é realizado depois do agrupamento, tendo como entrada o conjunto de dados X e os representantes $\mu_1, \mu_2, \dots, \mu_N$ de cada um dos N grupos determinados pela partição $C = \{C_1, C_2, \dots, C_N\}$. Os índices apresentados a seguir são definidos considerando $d(.,.)$ a medida de similaridade, m o número de elementos do grupo C_k e n o número de elementos que não estão no grupo C_k .

- ❑ Coesão : calcula o grau de similaridade entre os elementos de um mesmo grupo, (KUNZ; BLACK, 1995). Quanto maior a similaridade entre os elementos de um

mesmo grupo, maior a coesão desse grupo. Considerando x_i e $x_j \in C_k$, a coesão pode ser calculada por :

$$Coesao(C_k) = \frac{\sum_{i>j} d(x_i, x_j)}{\frac{m(m-1)}{2}}. \quad (1)$$

- Separação: mede a similaridade média de pares de elementos, formados por um elemento pertence a C_k e um outro externo ao grupo. Sejam $x_i \in C_k$ e $y_j \notin C_k$, a separação é dada pela equação (2), (KUNZ; BLACK, 1995).

$$Separacao(C_k) = \frac{\sum_{i,j} d(x_i, y_j)}{mn}. \quad (2)$$

- Silhueta: Seja $a(i)$ a média da dissimilaridade do i-ésimo objeto do grupo a todos os elementos do mesmo grupo e $b(i)$ a média da dissimilaridade do i-ésimo objeto ao grupo vizinho mais próximo. A silhueta SC é dada pela equação (3), em que N é o total de grupos (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2001).

$$SC = \frac{1}{N} \sum_{i=1}^N s(i); \quad s(i) = \frac{a(i) - b(i)}{\max\{a(i), b(i)\}} \quad (3)$$

Se $a(i)$ e $b(i)$ forem calculados como a distância do i-ésimo objeto ao centróide do grupo em questão temos uma Silhueta Simplificada. A Figura 4 ilustra as medidas $a(i)$ e $b(i)$, calculados como a distância do i-ésimo objeto ao centróide.

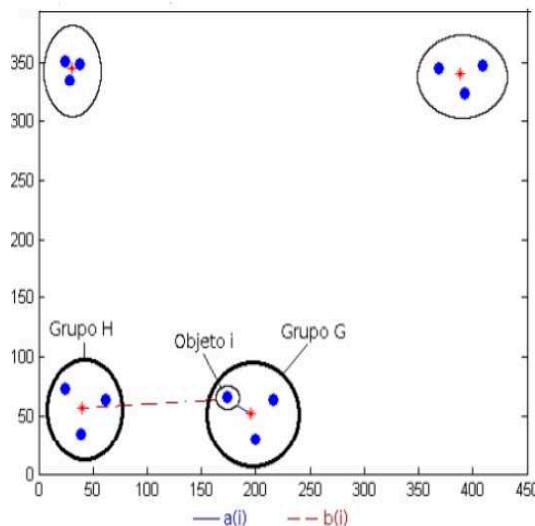


Figura 4 – Exemplo das medidas $a(i)$ e $b(i)$ para o cálculo da Silhueta.

Fonte: (VENDRAMIN; CAMPELLO; HRUSCHKA, 2010).

- *Coeficiente de Correlação Cofenética:* Este índice é utilizado na avaliação de agrupamentos hierárquicos. A partir da partição hierárquica tem-se matriz cofenética P_c cujos elementos são os valores das divergências mínimas que atenderam ao critério de fusão ou divisão dos grupos. Assim o elemento $P_c(i, j)$ representa a dissimilaridade

entre os dados x_i e x_j agrupados pela primeira vez. O Coeficiente de Correlação Cofenética mede a dissimilaridade entre a matriz original $P = (d_{ij})$ e a matriz obtida após a construção da hierarquia, $P_c = (c_{ij})$, é dado por:

$$CPCC = \frac{\frac{1}{M} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij} c_{ij} - \mu_P \mu_c}{\sqrt{[\frac{1}{M} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 - \mu_P^2][\frac{1}{M} \sum_{i=1}^{N-1} \sum_{j=i+1}^N c_{ij}^2 - \mu_c^2]}} \quad (4)$$

em que $M = N(N - 1)/2$, μ_P e μ_c as médias dos elementos das matrizes $P = (d_{ij})$ e $P_c = (c_{ij})$, respectivamente, definidas por:

$$\mu_P = \frac{1}{M} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij} \text{ e } \mu_c = \frac{1}{M} \sum_{i=1}^{N-1} \sum_{j=i+1}^N c_{ij}.$$

Note que o coeficiente de correlação cofenética é o coeficiente de Correlação de Person entre a matriz de dissimilaridade original e a obtida após a construção da hierarquia, P_c . Tem-se que $CPCC \in [-1, 1]$, quanto mais próximo de 1 maior é a similaridade entre as matrizes, (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2001).

2.3 Segmentação hierárquica

A partição hierárquica é uma representação multi-escala das possíveis segmentações da imagem a qual é frequentemente representada por uma estrutura de árvore. Na Figura 5, um exemplo dessa estrutura pode ser observada.

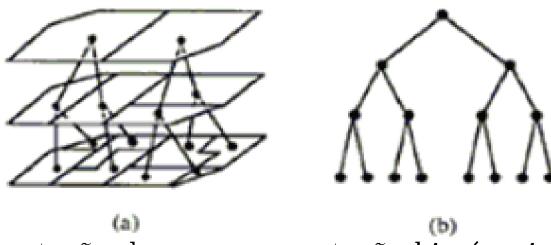


Figura 5 – (a) Representação de uma segmentação hierárquica. (b) Estrutura de Árvore da partição hierárquica.

Fonte: (BEAULIEU; GOLDBERG, 1989).

A segmentação hierárquica é um conjunto de partições em diferentes níveis de detalhes. As partições com menos detalhes podem ser produzidas a partir de fusões de regiões com mais riqueza de detalhes. Portanto, as segmentações mais finas estão aninhadas com aquelas em níveis mais grosseiros. A Figura 6 apresenta segmentações de uma mesma imagem utilizando partições representadas com diferentes escalas de refinamento.

Para obter a partição hierárquica, constrói-se, inicialmente, o Grafo de Regiões Adjacentes (RAG) dos pixels da imagem. A Figura 7 ilustra a partição mais refinada da segmentação hierárquica obtida associando cada nó do RAG a uma folha da estrutura de árvore. Para cada aresta do grafo associa-se um valor mínimo α o qual indica o custo

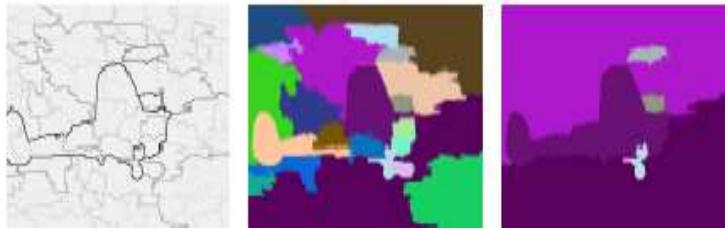


Figura 6 – Segmentações de uma imagem utilizando diferentes níveis de detalhes.
Fonte: (KLAVA B., 2010).

de associação. Depois fundem-se as regiões iterativamente seguindo o critério de fusão, geralmente esse critério implica a minimização do custo de fusão. Cada vez que um par de regiões é unido, um novo nó é criado na árvore. Este novo nó é unido ao par de nós, que agora são chamados nós filhos, e um novo valor α é atribuído ao “nó pai”. O processo finaliza quando existir somente uma região no RAG.

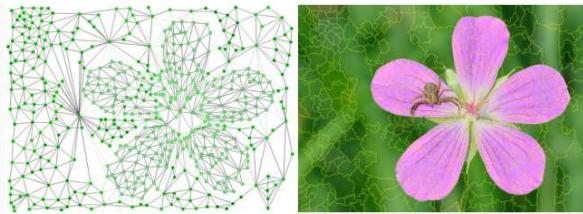


Figura 7 – Grafo de Regiões Adjacentes (RAG).
Fonte: (KLAVA B., 2010).

Com a estrutura hierárquica construída, o problema de segmentação passa a ser localizar a partição desejada na hierarquia, o que pode ser feito, como proposto em (KLAVA B., 2010), por meio das seguintes operações:

- Escolha do número de regiões, por meio da seleção de um limiar adequado para fusão.
- Re-segmentação de uma região, por meio da supressão de uma aresta de peso máximo na árvore correspondente a região em questão.

As operações de supressão de arestas na árvore geradora representa a partição e podem ser vistas como cortes na estrutura hierárquica, conforme ilustrado na Figura 8.

- a seleção do limiar para fusão da partição mais fina determina um corte em um mesmo nível.
- a re-segmentação de uma região determina um corte em um nível abaixo do nível atual da região selecionada.
- a fusão de regiões determina um corte em um nível acima do atual da região selecionada.

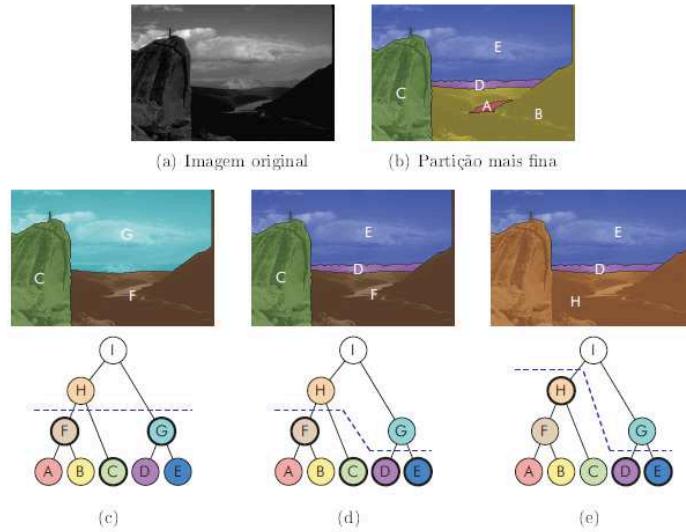


Figura 8 – Localização das partições.

Fonte:(KLAVA B., 2010).

Um modelo clássico utilizado na segmentação de uma imagem I é o funcional de Mumford-Shah (MUMFORD; SHAH, 1989). O procedimento de segmentação que utiliza este modelo variacional particiona I em conjuntos X_i obtidos pela aglomeração de regiões, considerando tanto a informação da região quanto o comprimento de suas fronteiras.

Koepfler (KOEPFLER; LOPEZ; MOREL, 1994) mostrou que é possível elaborar um algoritmo de segmentação via Mumford-Shah de complexidade $O(n \log n)$, em que n é o número de pixels da imagem. Assim, considerando $I = \{x_i\}$ o conjunto dos pixels de uma dada imagem, a segmentação é obtida minimizando o funcional:

$$E(u, K) = \beta \int \int_{\Omega} \|u - I\|_2^2 dx dy + \alpha \int \int_{\Omega/K} \|\nabla u\|_2^2 dx dy + \lambda l(K). \quad (5)$$

onde:

- ❑ $E(u, K)$ é a energia do funcional em função da imagem segmentada u e de suas fronteiras entre as regiões K ;
- ❑ K denota as fronteiras entre as regiões;
- ❑ $l(K)$ é comprimento das fronteiras entre as regiões;
- ❑ u denota a imagem segmentada;
- ❑ λ é parâmetro de escala;
- ❑ α e β são parâmetros de平衡amento dos termos da equação.

O processo de segmentação utilizando o funcional de Mumford-Shah pode ser visto como um problema de minimização dos termos da equação (5), os quais podem ser interpretados da seguinte forma:

- $\int \int_{\Omega} \|u - I\|_2^2 dx dy$ - estabelece a similaridade entre a segmentação u e a imagem I . Observe que quanto mais próximo de I a segmentação u estiver menor será a contribuição deste termo para a energia;
- $\int \int_{\Omega/K} \|\nabla u\|_2^2 dx dy$ - mede a variação de u dentro de cada região Ω_i em que $\Omega = \bigcup \Omega_i$. Desta forma, quanto mais constante for u dentro de cada região menor é a sua variação e menor o valor deste termo.
- $l(K)$ - calcula o comprimento das fronteiras. Quanto maior o comprimento das fronteiras, maior será a contribuição deste termo para a energia. Tornar mínima a contribuição deste termo inibe a super segmentação, gerando bordas econômicas e regulares.

O funcional de Mumford-Shah pode ser simplificado considerando a função u em cada região com valor constante igual a média dos valores desta região, assim não existe variação nas regiões e o termo $\int \int_{\Omega/K} \|\nabla u\|_2^2 dx dy$ é nulo. Tem-se então a equação (6) denominada forma simplificada do funcional de Mumford-Shah. Observe que este funcional opera em função de um fator de similaridade entre as regiões adjacentes comparadas e da fronteira K que divide estas regiões.

$$E(u, K) = \beta \int \int_{\Omega} \|u - I\|_2^2 dx dy + \lambda l(K). \quad (6)$$

Para decidir se duas regiões serão fundidas ou não, calcula-se a diferença entre a energia obtida considerando a fusão das regiões com a energia gerada quando estas permanecem separadas. As duas regiões adjacentes que levarem ao maior decréscimo de energia serão fundidas. Nesta perspectiva e utilizando a forma simplificada do funcional de Mumford-Shah o critério de fusão $\Delta(R_1, R_2)$ de duas regiões R_1 e R_2 pode ser discretizado pela equação:

$$\Delta(R_1, R_2) = \frac{|R_1||R_2|}{|R_1| + |R_2|} \|\mu_1 - \mu_2\|_2^2 - \lambda l(K); \quad (7)$$

onde $|R_i|$ e μ_i representam, respectivamente, a área e a média da intensidade (ou cor) de cada pixel da região R_i .

2.3.1 Medidas de Avaliação

Os algoritmos de segmentação podem ser avaliados analiticamente ou empiricamente. Os métodos analíticos avaliam os algoritmos por meio da análise de seus princípios, exigências, complexidade e propriedades.(ZHANG et al., 2009). Por outro lado, os métodos empíricos julgam a qualidade da segmentação de uma imagem gerada pelo processo de segmentação, comparando-a com uma segmentação de referência da imagem, ou seja uma segmentação da mesma imagem obtida manualmente a qual é suposta como verdadeira.

Desta forma, podem-se utilizar as medidas de avaliação externa: Índice de Jaccard e Índice de Folkes, definidas na seção 2.2.3, considerando:

- ❑ a=como o número de pixels que pertencem à intercessão de R_i e S_i , chamados positivos;
- ❑ b=número de pixels que pertence a S_i que não pertencem a R_i , chamados falso-positivos;
- ❑ c = número de pixels de R_i que não pertencem a S_i , chamados falso-negativos;

onde R_i ($i = 1, 2, \dots,$) é o i-ésimo segmento de referência e S_i o segmento produzido pelo algoritmo de segmentação com maior intercessão com R_i . A Figura 9 ilustra as definições apresentadas acima.

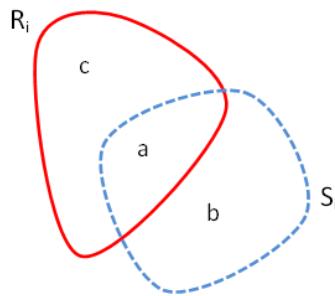


Figura 9 – Comparaçāo das regiões de referēcia R_i e segmentaçāo S_i .
Fonte: Imagem modificada de (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2001).

2.4 Recuperação de imagens baseada em conteúdo

Os sistemas de recuperação de imagens usam dois principais paradigmas: texto e conteúdo. Nas abordagens baseadas em texto, o processo de recuperação consiste em comparar os termos de uma consulta textual com anotações associadas, por exemplo, imagens representadas por palavras chave, (DATTA et al., 2008). A partir da comparação, o processo de recuperação retorna uma lista ordenada de imagens de acordo com a similaridade. Nesta abordagem, existem duas principais desvantagens: (1) a necessidade de um trabalho manual de anotações e (2) as anotações inadequadas devido à ambiguidade e incerteza das palavras utilizadas na anotação por pessoas. É importante destacar que o conteúdo de uma imagem é muito mais rico do que é possível descrever por meio de um conjunto de palavras (DATTA et al., 2008).

Os sistemas de recuperação de imagens baseados em conteúdo (CBIR - *Content-Based Image Retrieval*) surgem da tentativa de superar as desvantagens dos sistemas de recuperação de imagens baseados em texto (SCHWANDER; NIELSEN, 2010). Consistem

basicamente em caracterizar as imagens e medir a similaridade entre os vetores de características que representam essas imagens.

Os métodos de extração de características de imagens utilizam descritores, os quais representam a imagem por meio de um vetor que representa as propriedades visuais da imagem, tal como textura, forma e cor. A similaridade é obtida por meio da comparação dos descritores da imagem de consulta com os descritores das imagens da base de interesse.

Vários descritores baseados em cor, textura e forma têm sido propostos na literatura. Entretanto estes podem ser sensíveis às variações geradas por escala, rotação, mudanças na perspectiva, iluminação, níveis de contraste ou oclusão parcial. Alguns desses descritores são:

- SIFT (*Scale-Invariant Feature Transform*): descritor invariante a cor, transformações por escala rigidez e pequenas distorções (LOWE, 2004);
- BoVW (*Bag-of-Visual-Words*): técnica utilizada para representar e recuperar imagens estáticas, inclui o descritor SIFT na primeira etapa do seu processo para extrair as características das imagens do banco de dados. Em seguida, o conjunto de todos os SIFT são quantizados em K grupos, formando o dicionário. O centro de cada grupo é definido como uma palavra visual. Finalmente os descritores de cada SIFT são mapeado para a palavra visual mais próxima do dicionário (CSURKA et al., 2004);
- BSM (*Binary Salience Mape*) e FISM (*Fuzzy Image Descriptor Based on Saliency Map*): métodos que incluem a informação espacial do BoVW, mas diferem na forma utilizada na construção do histograma de palavras visuais. No BSM, utiliza-se o mapa de saliência (MS) enquanto que no FISM esta etapa é realizada por funções de pertinência *fuzzy*, classificando assim as palavras visuais de forma ponderada (NAKAMOTO; TORIU, 2011);
- SPM (*Spatial Pyramids Matching*): técnica utilizada para incorporar as informações espaciais globais e locais de uma imagem dentro de um vetor de característica (KRISTO; CHUA, 2013).

A Figura 10 demonstra o fluxograma de funcionamento do método BoVW que pode ser combinado com métodos de atenção visual (BSM e FISM) ou com SPM. Esta técnica tem sido foco de diversas pesquisas na área de visão computacional. Como exemplo, podemos citar os trabalhos de Dong (DONG; GUO; FU, 2014), Soares (SOARES; SILVA; GULIATO, 2012) e Liu (LIU et al., 2011).

O funcionamento do BoVW é constituído de três etapas:

- **Extração de característica:** consiste em identificar pontos locais de interesse em um conjunto de imagem representando-os por vetores de dimensão elevada, tais

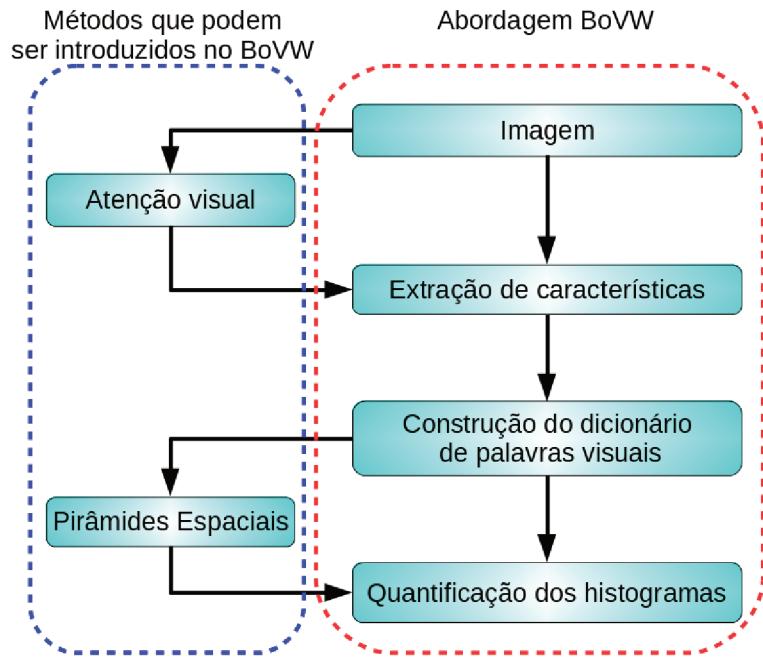


Figura 10 – Fluxograma da abordagem BoVW combinada com os métodos de atenção visual e pirâmides espaciais.

Fonte:(ROCHA, 2015)

como os descritores SIFT. Esses vetores são utilizados para construir o dicionário de palavras visuais;

- **Construção do dicionário de palavras visuais:** são utilizadas técnicas de agrupamento, tal como o *k-means*, para gerar os vocabulários. Nesse momento, os centróides de cada agrupamento são considerados como palavra visual e o conjunto dessas palavras formam o vocabulário também chamado de *codebook* ou dicionário;
- **Construção dos histogramas:** as características de cada imagem são mapeadas para a palavra visual mais próxima, obtendo assim um histograma de palavras visuais associado a cada imagem do banco de dados. O histograma resultante é conhecido como BoVW e sua dimensão está vinculada ao tamanho do dicionário.

A Figura 11 ilustra todo o processo para obtenção do dicionário de palavras visuais e para a descrição das imagens via histograma de frequência.

2.4.1 Medidas de avaliação

As duas medidas frequentemente utilizadas no contexto de recuperação de imagens são: a precisão e a revocação. A precisão (P) é a fração de documentos recuperados que são relevantes. A revocação (R) é a fração de documentos relevantes que são recuperados. Estas noções podem ficar mais claras examinando a seguinte tabela de contingência:

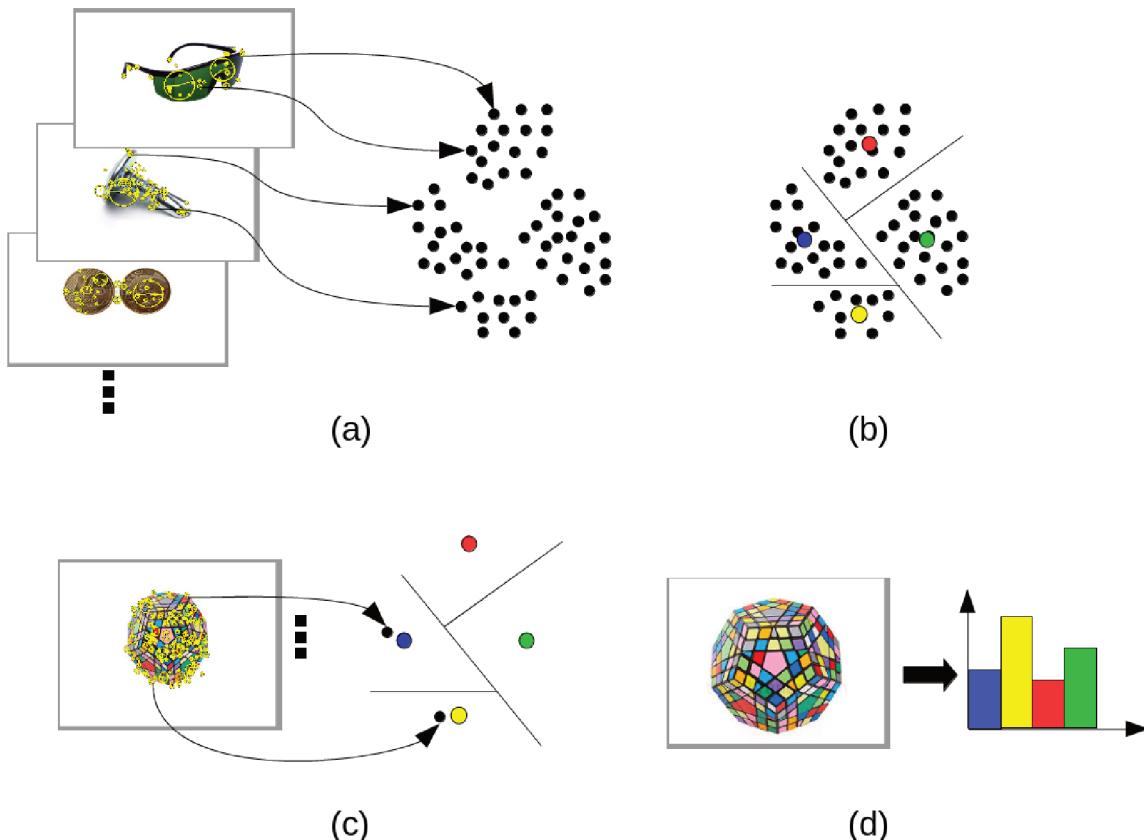


Figura 11 – Visão geral do *Bag of Visual Words*
Fonte: (ROCHA, 2015).

Tabela 1 – Tabela de Contigências

	Relevantes	Não relevantes
Recuperado	verdadeiros positivos (vp)	falsos positivos (fp)
Não recuperado	falsos negativos (fn)	verdadeiros negativos (vn)

Fonte: (POWERS, 2011)

A precisão e revocação são dadas por:

$$\begin{aligned}
 P &= \frac{vp}{(vp + fp)}; \\
 R &= \frac{vp}{(vp + fn)}.
 \end{aligned} \tag{8}$$

As medidas de precisão e revocação concentram-se na avaliação do retorno de verdadeiros positivos, considerando a porcentagem de documentos relevantes que são encontrados e a quantidade de falsos positivos retornados. Observe que em um bom sistema, a precisão geralmente diminui à medida que o número de documentos recuperados aumenta (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2001).

2.4.2 Precisão em k

Precisão em k (*precision at k*) é um fator de medida em todos os níveis de revocação. Para muitas aplicações é interessante avaliar a quantidade dos k primeiros resultados corretos exibidos. Assim faz-se necessário medir a precisão fixa de resultados recuperados, tal como 10 ou 30 documentos. Assim “*precision at k*”, em que k define a posição fixa da precisão, descreve a precisão obtida até uma quantidade fixa de resultados.

2.4.3 Mean Average Precision

A medida *Average Precision* (AP) é dada pela média dos valores das precisões obtidas pelo conjunto de *top k* documentos existentes após cada documento relevante ser recuperado. Isto é, para uma única consulta $q_j \in Q$, AP é a média das precisões computadas no ponto de cada item recuperado corretamente, $\{d_1, \dots, d_{m_j}\}$, na lista ordenada. A média dos valores *Average Precision* (AP) calculada para um conjunto de consultas Q é denominada *Mean Average Precision* e dada por:

$$MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} \text{precisão}(R_{jk}) \quad (9)$$

em que R_{jk} é o conjunto de resultados ranqueados, iniciando dos melhores até chegar ao item d_k .

O valor MAP para uma coleção de teste é a média aritmética dos valores das precisões média para uma única informação da precisão. Isto tem o efeito de ponderação equivalente para cada informação, mesmo que muitos documentos sejam relevantes para algumas consultas, ao passo que, poucos são relevantes para outras.

2.5 Registro variacional de imagens

O registro é uma técnica de processamento de imagens que tem como objetivo determinar uma transformação capaz de alinhar os pontos de uma imagem com pontos de outra que possua objetos semelhantes. O problema típico de registro ocorre quando duas imagens apresentam essencialmente o mesmo objeto, entretanto este encontra-se em posição diferente nas imagens. Assim, as imagens são espacialmente não alinhadas ou não registradas, ou seja, não há correspondência espacial direta entre elas. Normalmente, esse problema ocorre quando as imagens são tomadas a partir de perspectivas, tempo ou dispositivos diferentes (FISCHER; MODERSITZKI, 2004; BROWN, 1992).

Desta forma, dadas duas imagens de um mesmo objeto, uma sendo referida como imagem de referência S e o outra como modelo T , definidos em Ω_S e $\Omega_T \subset R^d$, respectivamente, procura-se um transformação vetorial $\phi(u)(.) : \Omega_S \rightarrow \Omega_T$ tal que:

$$\phi(u)(x) \mapsto x + u(x)$$

em que

$$u : \Omega_S \rightarrow \Omega_T,$$

$$u : x \mapsto u(x) = (u_1(x), u_2(x), \dots, u_d(x)),$$

de forma que o modelo T e a imagem de referência S transformada, $S \circ \phi(u(x)) = S(x + u(x))$ sejam similares, segundo alguma medida.

Como, em geral, as imagens referência e modelo são praticamente sobrepostas é comum considerar $\Omega_S = \Omega_T = \Omega$. Neste trabalho considera-se aplicações com $d = 2$. Tratando-se do espaço \mathbb{R}^2 tem-se que S é $S(x, y)$, T é $T(x, y)$ e $u = (u_1, u_2)$ com $u_1 = u(x, y)$ e $u_2 = u(x, y)$. Assim, para registrar $S(x, y)$ em $T(x, y)$ é preciso aplicar $S(x, y)$ em u para encontrar uma imagem $S^u(x', y')$ que seja o mais próxima possível a $T(x, y)$, para isso, considera-se $x' = x + u_1(x, y)$ e $y' = y + u_2(x, y)$.

Considerando a necessidade de alinhar espacialmente uma imagem S a uma imagem T , em que $S : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ e $T : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, o problema de registro se resume em encontrar um campo de deformação u tal que,

$$\min_u \mathfrak{D}(S^u, T), \quad (10)$$

em que $S^u(x) = S(x + u(x))$ e $\mathfrak{D}(S^u, T)$ é a medida de similaridade dada por $\mathfrak{D}(S^u, T) = \int_{\Omega} (S^u(x) - T(x))^2 dx$ (CHRISTENSEN; JOHNSON, 2001).

Várias medidas de similaridade têm sido usadas no processo de registro, entre elas destacam-se a diferença ao quadrado, o coeficiente de correlação, medidas baseadas em fluxo ótimo e medidas provenientes da teoria da informação, como “informação mútua”.

A soma das diferenças ao quadrado é um medida amplamente utilizada na literatura para comparar as intensidades de S e T , como por exemplo em (FISCHER; MODERSITZKI, 2004). Utilizando esta medida de similaridade o problema de registro pode ser formulado por:

$$\min_u \{E(u) = \frac{1}{2} \int_{\Omega} (S(x + u(x)) - T(x))^2 dx\}. \quad (11)$$

Entretanto, a solução obtida com o método de registro deve ser suave, pois, ao ser aplicada aos pontos de uma imagem, estes, não podem espalhar aleatoriamente. Logo, além de minimizar a distância, é importante garantir que o campo de deslocamento obtido seja suave. Para obtermos a solução u desejada é preciso adicionar um termo de suavização $\mathfrak{L}(u)$ que possibilita a regularidade da transformação ϕ , assim temos que,

$$\min_u \mathfrak{L}(u) + \mathfrak{D}(S^u, T),$$

em que \mathfrak{L} é o termo de suavização.

A escolha do termo de suavização depende da aplicação do método de registro, na literatura muitos autores propõem termos de suavização para trabalhar principalmente com imagens médicas, (ALVAREZ et al., 2007; CHRISTENSEN; JOHNSON, 2001; SAIKIT et al., 2008; CHEN; YEH; YIN, 2009; FISCHER; MODERSITZKI, 2004).

Assim, o problema de registo de imagem pode ser formulado como problema de minimização de um funcional de energia que compreende uma medida de similaridade e um termo regularizador. Ou seja, tomando $F(u) = \mathfrak{L}(u) + \mathfrak{D}(S^u, T)$, devemos determinar u tal que:

$$\min_u F(u).$$

Os autores em (FISCHER; MODERSITZKI, 2004) propuseram os termos de suavização $\|\Delta u\|^2$ e $\|\nabla u\|^2$ em que,

$$\begin{aligned}\|\nabla u\| &\triangleq \left(\sum_{j=1}^2 \|\nabla u_j\|^2 \right)^{1/2} \\ \|\nabla u_j\| &= \left(\int_{\Omega} |\nabla u_j(x)|^2 dx \right)^{1/2}.\end{aligned}$$

Equações similares são obtidas para $\|\Delta u\|$ quando o operador “ Δ ” é usado em vez de “ ∇ ”.

Assim o problema de registo de imagem pode ser formulado como problema de minimização dos seguintes funcionais:

$$F_c(S^u, T) = \|\Delta u\|^2 + \lambda \|S(x + u) - T(x)\|^2, \quad (12)$$

$$F_d(S^u, T) = \|\nabla u\|^2 + \lambda \|S(x + u) - T(x)\|^2, \quad (13)$$

em que λ é uma constante.

Após formulação variacional do problema do registro, o funcional pode ser minimizado mediante a aplicação de métodos variacionais. Inicialmente determina-se um sistema de equações de Euler-Lagrange que deve ser discretizado. Em seguida obtém-se a solução do sistema utilizando um método numérico apropriado, por exemplo, o método do gradiente descendente e métodos *multigrid* (CHUMCHOB, 2013). A solução numérica obtida é utilizada para discretizar as equações.

2.5.1 Equações de Euler-Lagrange

As equações de Euler-Lagrange permitem obter equações de movimento de um sistema para posteriormente representá-las em um espaço discreto. Na formalização de Lagrange um sistema é caracterizado por determinada função a qual depende das coordenadas generalizadas (y), das suas derivadas (\dot{y}) e também de tempo (x). Esta função é chamada Lagrangiana do sistema e é dada por $L(y, \dot{y}, x)$. Por meio do princípio da ação mínima, ou princípio de Hamilton, é possível mostrar um conjunto de equações diferenciais conhecidas por equações de Lagrange. Considerando as funções $y_i = (x_1, x_2, \dots, x_n)$, com $i = 1, \dots, m$,

$y_{ij} = \frac{dy_i}{dx_j}$ e o funcional:

$$J = \int_{x_1^{(1)}}^{x_1^{(2)}} dx_1 \int_{x_2^{(1)}}^{x_2^{(2)}} dx_2 \dots \int_{x_n^{(1)}}^{x_n^{(2)}} dx_n L(y_i, y_{ij}, x_j),$$

o sistema de equações de Euler-Lagrange é dado por:

$$\frac{\partial L}{\partial y_i} - \sum_{j=1}^n \frac{\partial}{\partial x_j} \left(\frac{\partial L}{\partial y_{ij}} \right) = 0 \quad (14)$$

em que $i = 1, \dots, m$.

Considerando $S(x)$ e $T(x)$ as imagens a serem registradas, e denotando $S^u(x) = S(x + u(x))$, a equação de Euler-Lagrange para (12) é:

$$-|\nabla u| \operatorname{Div} \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda(S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x})) \nabla S^u = 0, \quad (15)$$

e sua equação de fluxo:

$$\frac{\partial u}{\partial t} = |\nabla u| \operatorname{Div} \left(\frac{\nabla u}{|\nabla u|} \right) - \lambda(S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x})) \nabla S^u, \quad (16)$$

em que,

$$\begin{aligned} \nabla S^u &= \left(\frac{\partial S^u}{\partial u}, \frac{\partial S^u}{\partial v} \right), \\ |\nabla u| \operatorname{Div} \left(\frac{\nabla u}{|\nabla u|} \right) &= \frac{u_y^2 u_{yy} - 2u_x u_y u_{xy} + u_x^2 u_{xx}}{u_x^2 + u_y^2 + \beta}. \end{aligned}$$

De forma análoga temos que a equação de Euler-Lagrange que minimiza o funcional dado em (13) é:

$$-\Delta u + \lambda(S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x})) \nabla S^u = 0 \quad (17)$$

e sua equação de fluxo é:

$$\frac{\partial u}{\partial t} = \Delta u - \lambda(S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x})) \nabla S^u. \quad (18)$$

A partir da forma discretizada, obtém-se um algoritmo iterativo para a solução do problema de registro. A cada iteração é necessária a interpolação de $S(x + u^n)$ denotada por $\tilde{S}(x + u^n)$, pois, u é uma função contínua que possui valores reais e S é uma função discreta, computacionalmente a estrutura de dados utilizada para representar a imagem é uma matriz que não admite valores reais em suas coordenadas.

A seguir apresentamos, mais detalhadamente, os conceitos e métodos necessários a solução do problema de registro variacional, conforme descrito acima.

2.5.2 Método das Diferenças Finitas

Neste método, aproximações obtidas por diferenças finitas são utilizadas para solucionar as EDOs. As diferenças finitas podem ser atrasadas, centrais ou avançadas nas derivadas. Por meio da série de Taylor para a função y pode-se escrever

$$y_{n+1} = y_n + hy'_n(x) + \frac{h^2}{2!}y''_n(x) + \frac{h^3}{3!}y'''_n(x) + \dots$$

Utilizando uma expansão da série de Taylor até a primeira ordem tem-se:

$$y_{n+1} \approx y_n + hy'_n(x)$$

$$hy'_n(x) \approx y_{n+1} - y_n$$

Desta forma tem-se a diferença finita adiantada dada por

$$y'_n(x) \approx \frac{y_{n+1} - y_n}{h} \quad (19)$$

Outra fórmula de diferenças finitas pode ser obtida utilizando a seguinte expansão da série de Taylor até a primeira ordem:

$$y_{n-1} \approx y_n - hy'_n(x)$$

$$hy'_n(x) \approx y_n - y_{n-1}$$

$$y'_n(x) \approx \frac{y_n - y_{n-1}}{h}$$

A aproximação dada por

$$y'_n(x) \approx \frac{y_n - y_{n-1}}{h} \quad (20)$$

e denominada diferença finita atrasada.

As diferenças finitas centrais são obtidas somando as aproximações 19 de 20.

$$y'_n(x) \approx \frac{y_{n+1} - y_{n-1}}{2h} \quad (21)$$

A diferenças finitas atrasada, adiantada e central utilizadas como aproximações da segunda derivada são obtidas de modo análogo a partir da série de Taylor expandida até a segunda ordem. Assim, a segunda derivada de y pode ser aproximada utilizando a diferença central dada por:

$$y''_n(x) \approx \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2} \quad (22)$$

Sempre que possível, usam-se, para aproximar derivada, fórmulas em função das diferenças finitas centrais, tendo em vista que elas são mais precisas (SPERANDIO; MENDES; SILVA, 2003).

2.5.3 Interpolação

A interpolação determina uma aproximação do valor de uma função para um ponto no qual esta não está definida. Podem-se interpolar funções que definem imagem utilizando, por exemplo, os seguintes métodos:

Nearest Neighbors: método de interpolação que considera os quatro pixels mais próximos ao ponto a ser interpolado;

Dadas as coordenadas (x, y) de um ponto, em que x e y assumem valores reais, considerando u_1 a parte inteira de x e u_2 a parte inteira de y , os vizinhos mais próximos de (x, y) são: (u_1, u_2) , $(u_1 + 1, u_2 + 1)$, $(u_1, u_2 + 1)$, $(u_1 + 1, u_2)$. Isso é o mesmo que arredondar os valores de x e y , tal que a intensidade no pixel $(\text{round}(x), \text{round}(y))$ seja considerada como a intensidade em (x, y) , em que round é um função de arredondamento;

Bilinear: Semelhante ao método *Nearest Neighbors*, dadas as coordenadas (x, y) , esse método considera os quatro vizinhos próximos, no entanto, ao invés de um simples arredondamento, é feita uma soma ponderada entre aos quatro vizinhos para definir a nova intensidade de pixel. Matematicamente esse método pode ser expresso da seguinte forma:

$$\begin{aligned} I'(u_1, u_2) = & ((u_1 + 1 - x)(u_2 + 1 - y))I(u_1, u_2) \\ & + ((x - u_1)(u_2 + 1 - y))I(u_1 + 1, u_2) \\ & + ((u_1 + 1 - x)(y - u_2))I(u_1, u_2 + 1) \\ & + ((x - u_1)(y - u_2))I(u_1 + 1, u_2 + 1), \end{aligned}$$

onde os valores de I são as intensidades nos pontos (u_1, u_2) , $(u_1 + 1, u_2 + 1)$, $(u_1, u_2 + 1)$, $(u_1 + 1, u_2)$.

Esses métodos possuem um custo computacional baixo e facilmente implementados, sendo frequentemente utilizados nas técnicas de processamento de imagem.

2.5.4 Medidas de avaliação

Para mensurar a qualidade do registro entre duas imagens S e T , é preciso medir o quanto próxima de T está $S(x+u)$, duas medidas de performance podem ser utilizadas para isso: *Correlation Coefficient* (CC) e a *Root Mean Square Error* (RMSE). Essas medidas são definidas por:

$$CC(S, T) = \frac{\sum_{i=1}^N \sum_{j=1}^M (S_{i,j} - \bar{S})(T_{i,j} - \bar{T})}{\sqrt{\sum_{i=1}^N \sum_{j=1}^M (S_{i,j} - \bar{S})^2 \sum_{i=1}^N \sum_{j=1}^M (T_{i,j} - \bar{T})^2}}, \quad (23)$$

e

$$RMSE(S, T) = \sqrt{\frac{1}{|\Omega|} \sum_{i=1}^N \sum_{j=1}^M (S_{i,j} - T_{i,j})^2}, \quad (24)$$

em que $S_{i,j}$ e $T_{i,j}$ são a intensidade de pixel (i, j) para as imagens S e T , respectivamente e $|\Omega|$ é o número total de pixels. \bar{S} e \bar{T} são a média de intensidade de S e T , respectivamente.

CAPÍTULO 3

Divergências de Bregman e Total Bregman: Definição, domínio e Tratamentos

As divergências de Bregman (DB) têm sido, tradicionalmente, aplicadas à problemas de otimização e análise numérica, mas recentemente têm contribuído para avanços nas pesquisas envolvendo aprendizagem estatística (GUO; HONG; YANG, 2017), clusterização (LIU; BELKIN, 2016), e problemas inversos (BURGER, 2016).

O conceito da Divergência Total Bregman (DTB) é derivado do método de aproximação por mínimos quadrados total, utilizado para determinar a curva que melhor se aproxima de um conjunto de pontos. Observou-se que enquanto o método tradicional de mínimos quadrados depende dos sistemas de coordenadas, o método de mínimos quadrados total é invariante. A definição de divergência Total Bregman foi obtida a partir da divergência de Bregman, de forma análoga.

As divergências de Bregman e Total Bregman determinam a similaridade entre dois pares de “objetos” representados por números, vetores, matrizes ou funções densidade de probabilidades. Neste trabalho, as divergências são utilizadas para quantificar a dissimilaridade entre imagens caracterizadas por diferentes formas.

Com o intuito de detalhar a teoria utilizada são apresentados, a seguir, definições e propriedade das divergências de Bregman e Total Bregman. Ainda, considerando o domínio destas divergências são definidos e avaliados tratamentos que possibilitam a utilização destas no contexto da análise de imagens.

3.1 Divergência de Bregman

Definição: Dada uma função convexa $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ a correspondente divergência de Bregman entre x e $y \in \text{dom}(\varphi)$ é dada por:

$$d_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle \nabla \varphi(y), x - y \rangle. \quad (25)$$

$\nabla \varphi(y)$ é o vetor gradiente de φ em y .

Essa definição é, embora com pouco rigor matemático, exemplificada na Figura 12.

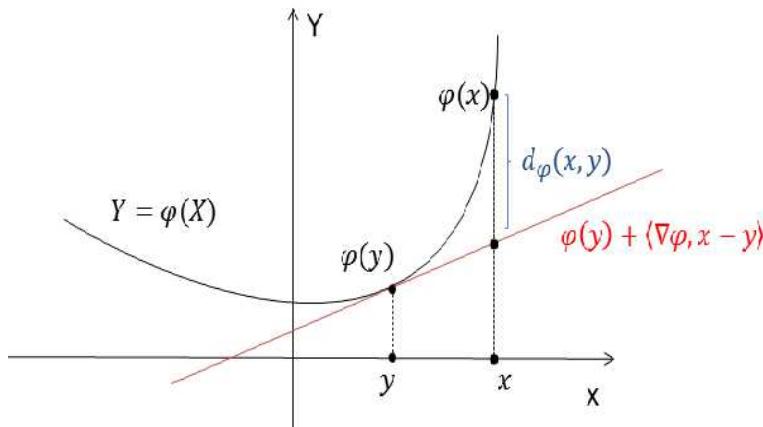


Figura 12 – Interpretação geométrica da divergência de Bregman no \mathbb{R}^2 .

Divergências entre x e y pertencentes a $X \subset \mathbb{R}^d$ tal que $d > 1$, podem ser facilmente construídas utilizando funções geradoras dadas pela soma das funções geradoras unidimensionais aplicadas em cada coordenada de $x \in \mathbb{R}^d$, isto é,

$$\varphi(x) = \sum_{i=1}^d \phi(x_i).$$

Desta forma, as divergências podem ser calculadas em cada coordenada utilizando as divergências obtidas a partir de funções geradoras definidas em subconjuntos de uma dimensão. As divergências assim obtidas são denominadas divergências Separáveis, (NI-ELSEN; PIRO; BARLAUD, 2009). Na Tabela 2 apresentamos DB obtidas considerando algumas funções geradoras unidimensionais

A Tabela 3 contém algumas funções convexas φ definidas em domínios de dimensão $d > 1$ com suas respectivas divergências de Bregman.

A seguir apresentam-se algumas propriedades importantes das divergências de Bregman, (BANERJEE et al., 2005b).

- **Não negatividade:** $d_\varphi(\mathbf{x}, \mathbf{y}) \geq 0$, $\forall \mathbf{x} \in \mathcal{S}$, $\mathbf{y} \in \text{ri}(\mathcal{S})$, e $d_\varphi(\mathbf{x}, \mathbf{y}) = 0$ se e somente se $\mathbf{x} = \mathbf{y}$.
- **Convexidade:** d_φ é sempre convexa no primeiro argumento, mas não necessariamente convexa no segundo argumento. A distância *Squared loss* e *KL-Divergence*

Tabela 2 – Funções Convexas unidimensionais e divergências de Bregman.

Domínio X	Função ϕ	$d_\phi(x, y)$	Divergência
\mathbb{R}	x^2	$(x - y)^2$	Squared loss
\mathbb{R}_+	$x^\alpha, \alpha \in \mathbb{R}, \alpha > 1$	$x^\alpha + (\alpha - 1)y^\alpha - \alpha xy^{\alpha-1}$	Norm-like
\mathbb{R}_+	$x \log x - x$	$x \log(\frac{x}{y}) - x + y$	Generalized I-divergence
\mathbb{R}_+^*	$-\log x$	$\frac{x}{y} - \log(\frac{x}{y}) - 1$	Itakura-Saito
$[0, 1]$	$x \log_2 x$	$x \log_2(\frac{x}{y})$	KL-Divergence
$[0, 1]$	$x \log x + (1 - x) \log(1 - x)$	$x \log(\frac{x}{y}) + (1 - x) \log(\frac{1-x}{1-y})$	Logistic loss
\mathbb{R}	$\log(1 + \exp x)$	$\log(\frac{1+\exp x}{1+\exp y}) + (x - y) \frac{\exp y}{1+\exp y}$	Dual logistic loss
$[-1, 1]$	$-\sqrt{1 - x^2}$	$\frac{1-xy}{\sqrt{1-y^2}} - \sqrt{1 - x^2}$	Hellinger-like

Tabela 3 – Divergências de Bregman Separáveis.

Domínios	$\varphi(\mathbf{x})$	$d_\varphi(\mathbf{x}, \mathbf{y})$	Divergência
\mathbb{R}^d	$\ \mathbf{x}\ ^2$	$\ \mathbf{x} - \mathbf{y}\ ^2$	Squared loss
\mathbb{R}^d	$\mathbf{x}^T A \mathbf{x}$	$(\mathbf{x} - \mathbf{y})^T A (\mathbf{x} - \mathbf{y})$	Mahalanobis
$d\text{-Simplex}$	$\sum_{j=1}^d x_j \log_2 x_j$	$\sum_{j=1}^d x_j \log_2(\frac{x_j}{y_j})$	KL-Divergence
\mathbb{R}_+^d	$\sum_{j=1}^d x_j \log x_j$	$\sum_{j=1}^d x_j \log(\frac{x_j}{y_j}) - \sum_{j=1}^d (x_j - y_j)$	Generalized I-Divergence (GID)

são exemplos de divergências de Bregman que são convexas em seus dois argumentos.

Porém um exemplo de divergência que não é convexa em y é a função estritamente convexa $\varphi(x) = x^3$, definida em \mathbb{R}_+ , dado por $d_\varphi(x, y) = x^3 - y^3 - 3(x - y)y^2$.

□ **Linearidade:** A divergência de Bregman é um operador linear, ou seja, $\forall \mathbf{x} \in \mathcal{S}, \mathbf{y} \in \text{ri}(\mathcal{S})$,

$$\begin{aligned} d_{\varphi_1 + \varphi_2}(\mathbf{x}, \mathbf{y}) &= d_{\varphi_1}(\mathbf{x}, \mathbf{y}) + d_{\varphi_2}(\mathbf{x}, \mathbf{y}), \\ d_{c\varphi}(\mathbf{x}, \mathbf{y}) &= cd_\varphi(\mathbf{x}, \mathbf{y}), \text{ (para } c \geq 0\text{)} . \end{aligned}$$

□ **Separação linear:** A localização de todos os pontos $\mathbf{x} \in \mathcal{S}$ que são equidistantes de dois pontos fixos $\mu_1, \mu_2 \in \text{ri}(\mathcal{S})$ é um hiperplano, ou seja, as partições induzidas pela divergência de Bregman têm separadores lineares dados por:

$$\begin{aligned} d_\varphi(\mathbf{x}, \mu_1) &= d_\varphi(\mathbf{x}, \mu_2) \\ \Rightarrow \varphi(\mathbf{x}) - \varphi(\mu_1) - \langle \mathbf{x} - \mu_1, \nabla \varphi(\mu_1) \rangle &= \varphi(\mathbf{x}) - \varphi(\mu_2) - \langle \mathbf{x} - \mu_2, \nabla \varphi(\mu_2) \rangle \\ \Rightarrow \langle \mathbf{x}, \nabla \varphi(\mu_2) - \nabla \varphi(\mu_1) \rangle &= (\varphi(\mu_1) - \varphi(\mu_2)) - (\langle \mu_1, \nabla \varphi(\mu_1) \rangle - \langle \mu_2, \nabla \varphi(\mu_2) \rangle) \end{aligned}$$

□ **Teorema de Pitágoras generalizado:** Para qualquer $\mathbf{x}_1 \in \mathcal{S}$ e $\mathbf{x}_2, \mathbf{x}_3 \in \text{ri}(\mathcal{S})$, a seguinte propriedade entre três pontos se estabelece:

$$d_\varphi(\mathbf{x}_1, \mathbf{x}_2) + d_\varphi(\mathbf{x}_2, \mathbf{x}_3) \leq d_\varphi(\mathbf{x}_1, \mathbf{x}_3).$$

Em (BANERJEE et al., 2005b), a divergência de Bregman é utilizada para medir a taxa de distorção de Shannon, introduzindo assim, o conceito de Informação de Bregman.

A partir deste conceito o problema de encontrar uma partição e seus respectivos representantes pode ser apresentado como um problema de minimização da Perda da Informação de Bregman. Nas próximas seções apresentam-se de forma resumida estes conceitos.

3.1.1 Informação de Bregman

Definição: Seja X uma variável aleatória com valores em $X = \{x_i\}_{i=1}^n \subset S \subset \mathbb{R}^n$ seguindo uma distribuição de probabilidade discreta $v = \{v_i\}$. Dada a divergência de Bregman d_φ a *Informação de Bregman* de X com relação a φ é dada por:

$$I(X) = \min_{s \in ri(S)} \sum_{i=1}^n v_i d_\varphi(x_i, s).$$

onde $ri(S)$ é o interior relativo de S , definido para todo conjunto convexo não vazio $S \subseteq \mathbb{R}^d$ por: $ri(S) := \{x \in S : \forall y \in S \exists \lambda > 1 : \lambda x + (1 - \lambda)y \in S\}$ (ROCKAFELLAR, 1996).

O vetor s que torna o funcional acima mínimo é chamado **Representante de Bregman de X**. Este representante não depende da escolha da divergência de Bregman e é o valor esperado para a variável X (BANERJEE et al., 2005b). Portanto tomando,

$$s = \mu = \sum_{i=1}^n v_i x_i \quad (26)$$

tem-se que,

$$I(X) = \sum_{i=1}^n v_i d_\varphi(x_i, \mu). \quad (27)$$

3.1.2 Perda da Informação de Bregman

Seja \mathbf{X} uma variável aleatória com valores em $X = \{x_i\}_{i=1}^n \subset S \subset \mathbb{R}^n$ e com medida de probabilidade v . Considere $\wp = \{R_h\}_{h=1}^k$ uma partição de X com seus respectivos representantes de Bregman. Se $M = \{\mu_h\}_{h=1}^k$ é o conjunto destes representantes e $p = \{p_h\}_{h=1}^k$ com $p_h = \sum_{x_i \in R_h} v_i$ uma medida de probabilidade em M , tem-se a variável induzida \mathbf{M} com valores em M e distribuição de probabilidade p . Assim \mathbf{R}_h é uma variável aleatória com valores em R_h e distribuição de probabilidade $\frac{v_i}{p_h}$ para $x_i \in R_h$.

Observe que $I_\varphi(\mathbf{X})$ é a “Informação de Bregman Total” enquanto $I_\varphi(\mathbf{M})$ é a “Informação de Bregman entre as regiões”. A diferença $L_\varphi(\mathbf{M}) = I_\varphi(\mathbf{X}) - I_\varphi(\mathbf{M})$ é denominada **Perda da Informação de Bregman** e representa a “Informação de Bregman dentro das regiões”. Aplicando a definição de Informação de Bregman dada em (27), segue que:

$$L_\varphi(\mathbf{M}) = \sum_{h=1}^k \sum_{i=1}^n v_i d_\varphi(x_i, \mu_h). \quad (28)$$

Note que $L_\varphi(\mathbf{M})$ mede a Perda de Informação ocorrida ao realizar a partição de X . Sendo assim, quanto menor a perda de informação de Bregman melhor será a partição realizada.

Se o número de regiões k , dado pela partição, for igual ao total de elementos do conjunto X , não existe perda de informação. Entretanto, se todos os elementos do conjunto X forem agrupados em um único conjunto, a perda da informação é dada pela informação de Bregman da variável \mathbf{X} , ou seja, $L_\varphi(\mathbf{M}) = I_\varphi(\mathbf{X})$ sendo $M = \{\mu\}$ e μ o representante de Bregman de \mathbf{X} .

Com base nas definições apresentadas anteriormente o problema de determinar uma partição para uma dada imagem $I = \{x_i\}_{i=1}^n$ pode ser formulado como um problema de otimização de $L_\varphi(\mathbf{M})$. O objetivo é determinar $M = \{\mu_h\}_{h=1}^k$ tal que $L_\varphi(\mathbf{M})$ seja mínimo.

3.1.3 Árvores de Bolas de Bregman (*BB-tree*)

Diversas estruturas de dados utilizadas em problemas de busca por proximidade são baseados na decomposição do espaço. Basicamente a ideia é dividir o espaço de busca em partes menores, por exemplo, retângulos, cubos, bolas e esferas. A partição do espaço é armazenada na forma de árvore que pode ser construída de duas maneiras: de cima para baixo ou de baixo para cima. No primeiro caso, o nó raiz corresponde à célula contendo todo o banco de dados, o qual é dividido de forma iterativa até que um critério de convergência seja satisfeito. No segundo caso, começa-se com um conjunto inicial de regiões, muitas vezes chamados de sementes, fundidos de forma iterativa, as regiões resultantes são representadas como um novo nó que é o pai das regiões mescladas (JAIN; MURTY; FLYNN, 1999).

Lawrence Cayton apresenta um algoritmo de busca por vizinhos mais próximos que não utiliza a desigualdade triangular e baseia-se na decomposição hierárquica do espaço em corpos convexos denominados bolas de Bregman (BB). Esta estrutura de dados é denominada *árvore de bolas de Bregman (BB-tree)*, (CAYTON, 2008).

Para definição da forma que representa cada parte, retângulo ou bolas, por exemplo, deve-se considerar a facilidade para estabelecer a divergência de um ponto ou um elemento de consulta, com relação a estas células. Em (CAYTON, 2008), as células utilizadas são bolas de Bregman definidas por:

$$B_\varphi(\mu, R) = \{x | d_\varphi(x, \mu) < R\}$$

em que d_φ é uma divergência de Bregman e μ o representante de Bregman definido em (26).

Considerando a base de dados $X = \{x_1, x_2, \dots, x_n\}$ a BB-tree é uma árvore binária na qual cada nó i está associado a um subconjunto $X_i \subset X$. Para todo nó i tem-se uma bola de Bregman com centro μ_i e raio R_i tal que $X_i \subset B_\varphi(\mu_i, R_i)$. Os nós, com exceção das folhas, têm nós filhos l e r . A Figura 13 ilustra a decomposição do espaço utilizando as BB-trees. A Figura 14 apresenta bolas abertas de Bregman, considerando as divergências: *Squared loss*, *Generalized I-divergence* (GID), *Itakura Saito* e *Logistic loss*.

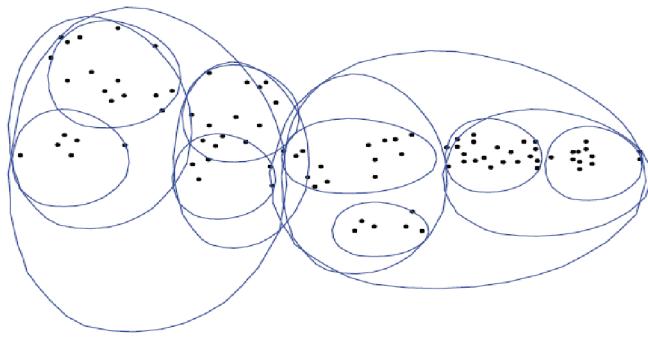


Figura 13 – Exemplo de Árvore de Bolas de Bregman
Fonte: (CAYTON, 2008).

As bolas foram obtidas considerando pontos x pertencentes ao conjunto $[0, 1] \times [0, 1] \subset \mathbb{R}^2$, $c = (0, 15, 0, 15)$ e raio $R = 0, 1$.

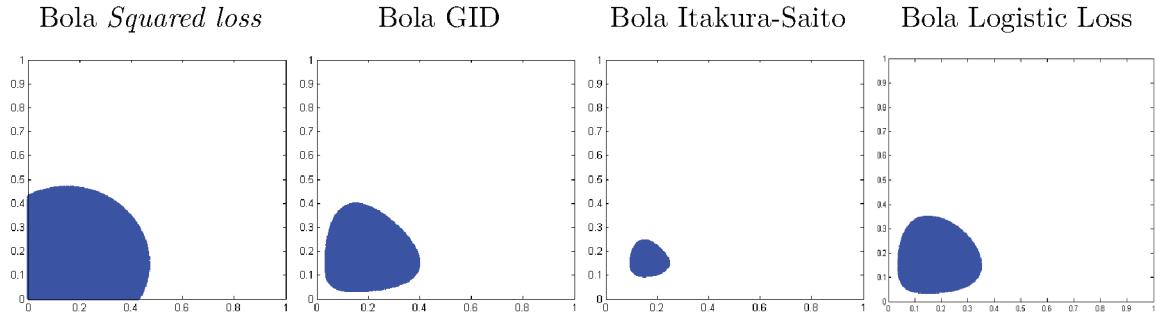


Figura 14 – Bolas de Bregman

Desta forma, a partir da escolha da DB, é possível particionar o espaço utilizando diferentes formas na representação de cada parte.

A Figura 15 ilustra a partição da região do plano, $[0, 1] \times [0, 1]$, utilizando as divergências: *Squared loss*, *Generalized I-divergence* e *Itakura Saito*. A partição do espaço foi realizada considerando bolas cujos centros c são dados pelos pontos que minimizam a divergência de Bregman entre os demais elementos de um grupo e eles (rigth-sided centroid).

Estes centros não dependem da escolha da divergência de Bregman e são os centróides dos conjuntos (BANERJEE et al., 2005a). Dado um conjunto $X = \{x_1, x_2, \dots, x_n\}$ o centróide, μ , é definido por:

$$c = \mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (29)$$

As diferentes divergências possibilitam obter agrupamentos considerando a topologia dos dados, já que dividem o espaço em regiões com formas diferentes.

Na Figura 16, apresentamos uma comparação dos resultados obtidos na clusterização de um conjunto de dados utilizando a divergência *Squared loss* e *Itakura Saito*. A Figura

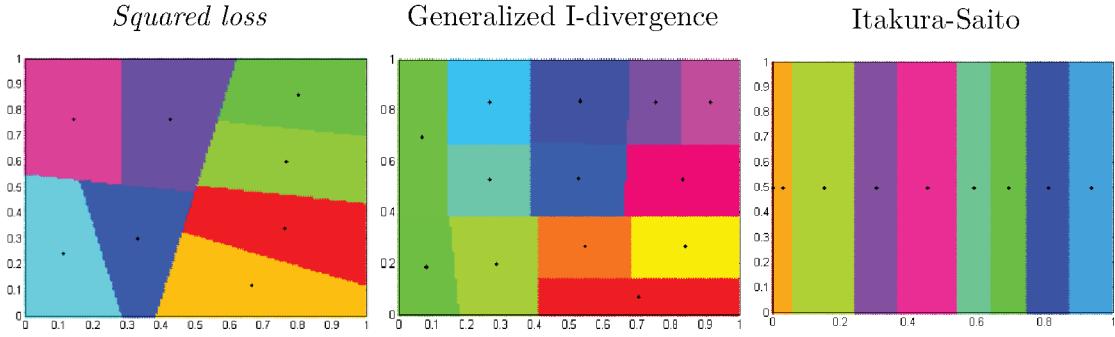
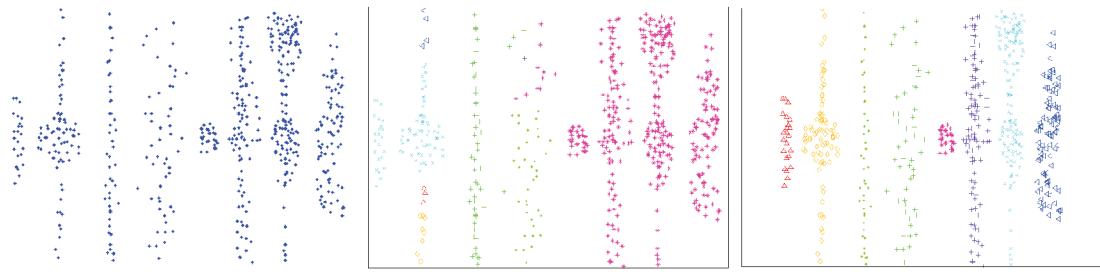


Figura 15 – Partição do Espaço utilizando diferentes DB

Figura 16 – (a)Conjunto de dados, (b) Clusters obtidos utilizando a divergência *Squared loss* (c) Clusters resultantes utilizando a divergência *Itakura Saito* .

16 (a) apresenta o conjunto de dados utilizado nesta comparação. Os dados sintéticos são definidos em 2 dimensões e contêm oito grupos irregulares com formas assimétricas e diferentes tamanhos

A topologia dos dados sugere que a partição do espaço utilizando Bolas *Itakura Saito* seja mais apropriada para determinar os agrupamentos. De fato, a divergência *Squared loss* quando empregada na clusterização deste conjunto apresenta resultados com erros e não foi capaz de definir os oitos grupos, Figura 16 (b). Enquanto que o método utilizando a divergência *Itakura Saito* determinou com eficácia os oito clusters do conjunto de dados, Figura 16 (c).

3.2 Divergência Total Bregman

Definição: Dada uma função convexa φ , denominada função geradora, a correspondente divergência Total Bregman entre x e $y \in \text{dom}(\varphi)$ é dada por,

$$tD_\varphi(x, y) = \frac{d_\varphi(x, y)}{\sqrt{1 + \|\nabla\varphi(y)\|^2}} \quad (30)$$

em que $d_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x - y \rangle$ é a divergência de Bregman e $\nabla\varphi(y)$ o vetor gradiente de φ em y .

Observe que a divergência Total Bregman é igual a divergência de Bregman multiplicada por um termo denominado *fator conformal* ρ_φ (NOCK; NIELSEN; AMARI, 2016):

$$tD_\varphi(x, y) = \frac{d_\varphi(x, y)}{\sqrt{1 + \|\nabla\varphi(y)\|^2}} = \rho_\varphi(y)d_\varphi(x, y),$$

$$\rho_\varphi(y) = \frac{1}{\sqrt{1 + \|\nabla\varphi(y)\|^2}}.$$

O fator conformal tem um papel regularizador no cálculo dos centróides pois é inversamente proporcional ao módulo do gradiente das funções convexas utilizadas. Essa propriedade faz com que dados “ruidosos” e “outliers” tenham menor peso que dados normais, visto que o gradiente dos dados ruidosos têm maior módulo do que os dados normais, (NIELSEN; NOCK, 2015).

Geometricamente, a divergência Total Bregman mede a projeção ortogonal entre o valor da função φ no primeiro argumento x e o hiperplano tangente a φ em $(y, \varphi(y))$, Figura 17. Observe que, diferentemente da divergência de Bregman, esta medida permanece inalterada com a rotação do sistema de coordenadas. A Tabela 4 apresenta divergências

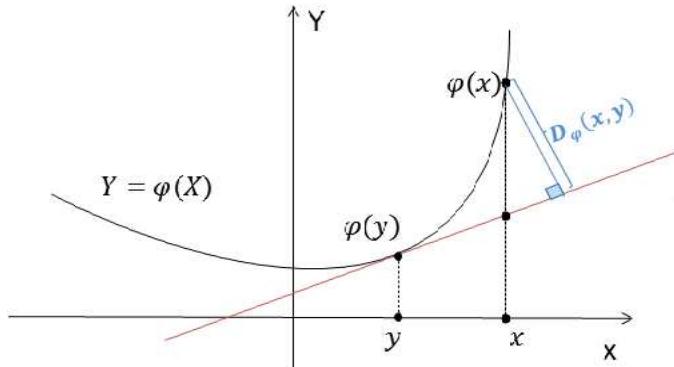


Figura 17 – Interpretação geométrica da divergência Total Bregman.

Total Bregman obtidas a partir de funções geradoras, $\phi(x)$, definidas em subconjuntos de uma dimensão.

A diferença de comportamento entre as divergências de Bregman e Total Bregman pode ser visualizada nos gráficos apresentados na Figura 18. Os gráficos ilustram a divergência entre x , y no domínio $X = [0, 1]$. Observamos que as DTB têm imagens que variam no intervalo $[0, 1]$, enquanto que algumas DB apresentam conjuntos de imagens com limites superiores mais altos.

Em muitas aplicações de visão computacional tais como recuperação de imagem, clusterização e classificação, é comum procurar um representante ou modelo para um conjunto de objetos que têm características semelhantes. Este representante é identificado como centro do conjunto. É desejável que os centros possam ser estabelecidos com baixo custo computacional. Os centros definidos utilizando a divergência Total Bregman são denominados ***t*-centros**, (LIU et al., 2012).

Tabela 4 – Divergências Total Bregman entre pontos unidimensionais (escalares) utilizadas para obter divergências Total Bregman Separáveis.

Domínio X	Função ϕ	Fator Conformal	$tD_\phi(x, y)$
\mathbb{R}	x^2	$\frac{1}{\sqrt{1+4x^2}}$	Total <i>Squared loss</i> $\frac{(x-y)^2}{\sqrt{1+4y^2}}$
\mathbb{R}_+	$x^\alpha, \alpha \in \mathbb{R}, \alpha > 1$	$\frac{1}{\sqrt{1+(\alpha x^{\alpha-1})^2}}$	Total <i>Norm-like</i> $\frac{x^\alpha + (\alpha-1)y^\alpha - \alpha xy^{\alpha-1}}{\sqrt{1+(\alpha y^{\alpha-1})^2}}$
\mathbb{R}_+	$x \log x - x$	$\frac{1}{\sqrt{1+(\log x)^2}}$	Total I-divergence $\frac{x \log(\frac{x}{y}) - x + y}{\sqrt{1+(\log y)^2}}$
\mathbb{R}	$\exp x$	$\frac{1}{\sqrt{1+(\exp x)^2}}$	Total <i>Exponencial loss</i> $\frac{\exp x - (x-y+1) \exp y}{\sqrt{1+(\exp y)^2}}$
\mathbb{R}_+^*	$-\log x$	$\frac{1}{\sqrt{1+(\frac{1}{x})^2}}$	Total Itakura-Saito $\frac{\frac{x}{y} - \log(\frac{x}{y}) - 1}{\sqrt{1+(\frac{1}{y})^2}}$
$(0, 1)$	$x \log x + (1-x) \log(1-x)$	$\frac{1}{\sqrt{1+(\log(\frac{x}{1-x}))^2}}$	Total <i>Logistic loss</i> $\frac{x \log(\frac{x}{y}) + (1-x) \log(\frac{1-x}{1-y})}{\sqrt{1+(\log(\frac{y}{1-y}))^2}}$
\mathbb{R}	$\log(1 + \exp x)$	$\frac{1}{\sqrt{1+(\log(\frac{\exp x}{1+\exp x}))^2}}$	Total <i>Dual logistic loss</i> $\frac{\log(\frac{1+\exp x}{1+\exp y}) + (x-y) \frac{\exp y}{1+\exp y}}{\sqrt{1+(\log(\frac{\exp y}{1+\exp y}))^2}}$
$[-1, 1]$	$-\sqrt{1-x^2}$	$\frac{1}{\sqrt{1+\frac{x^2}{1-x^2}}}$	Total <i>Hellinger-like</i> $\frac{\frac{1-xy}{\sqrt{1-y^2}} - \sqrt{1-x^2}}{\sqrt{1+\frac{y^2}{1-y^2}}}$

3.2.1 t-Centro

Dado um conjunto $X = \{x_1, x_2, \dots, x_n\}$ o t-centro é dado por

$$\bar{x} = \frac{(\sum w_i x_i)}{\sum w_i}, \quad w_i = \frac{1}{(1 + \|\nabla \varphi(x_i)\|^2)^{\frac{1}{2}}}.$$

Considerando, por exemplo, a função convexa $\varphi = \|x\|^2$ o t- centro é dado por

$$\bar{x} = \frac{(\sum w_i x_i)}{\sum w_i}, \quad w_i = \frac{1}{\sqrt{1+4x^2}}.$$

O t-centro pode ser visto como o representante de um conjunto que minimiza a divergência Total Bregman (DTB) entre ele e os demais elementos do grupo (left-sided centroid) (LIU et al., 2012) e é definido como a média ponderada cujos pesos atribuídos são inversamente proporcionais ao módulo do gradiente das funções convexas utilizadas.

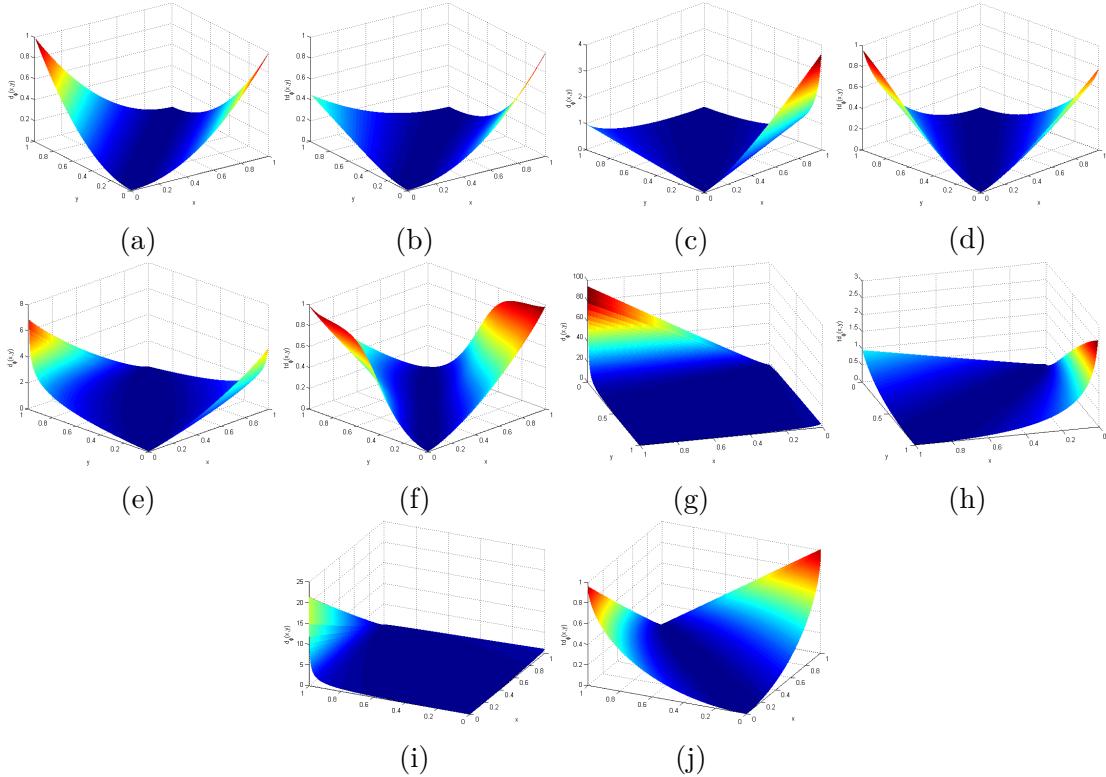


Figura 18 – Diferença de comportamento entre DB e DTB: (a) *Squared loss*, (b) *Total Squared loss*, (c) *Generalized I-divergence*, (d) *Total Generalized I-divergence*, (e) *Logistic loss*, (f) *Total Logistic loss*, (g) *Itakura-Saito* (h) *Total Itakura-Saito*, (i) *Hellinger-like* and (j) *Total Hellinger-like*.

Essa propriedade faz com que dados “ruidosos” e “outliers” tenham menor peso que dados normais, visto que o gradiente dos dados ruidosos têm maior módulo do que os dados normais. Dessa forma, pode-se concluir que o t -centro é robusto com relação a dados ruidosos e *outliers*. Devido a sua forma analítica, esses representantes podem ser facilmente obtidos com eficácia e com custo e tempo computacional reduzidos, (LIU, 2011).

As propriedades do t-centro justificam a análise da DTB nas diversas aplicações da processamento de imagens, já que este é definido a partir destas divergências.

3.3 Tratamento dos dados para a aplicação das Divergências de Bregman e Total Bregman

As divergências de Bregman e Total Bregman, são medidas adequadas para estabelecer a similaridade entre dados que caracterizam as imagens. Entretanto a aplicação destas divergências exige que tratamentos sejam realizados, visto que é necessário, devido às características dos dados, garantir o domínio de aplicação das respectivas divergências. Observe por exemplo, a divergência de Kullback Leibler (KL), dada por:

$$\sum_{j=1}^d x_j \log_2\left(\frac{x_j}{y_j}\right).$$

em que \mathbf{x} e \mathbf{y} pertencem ao conjunto d -Simplex,

$$\Delta^d = \{\mathbf{x} = (x_1, x_2, \dots, x_d); 0 \leq x_j \leq 1 \quad e \quad \sum_{j=1}^d x_j = 1\}.$$

Suponha que, para algum j , $x_j = 0$ e $y_j \neq 0$, a parcela $0\log_2(0/y_j)$ seja considerada nula, visto que o limite $\lim_{x_j \rightarrow 0} x_j \log_2 x_j = 0$. Por outro lado se $x_j \neq 0$ e $y_j = 0$, pode-se assumir que $x_j \log_2(x_j/0) = \infty$ essa convenção justifica-se por continuidade, (COVER; THOMAS, 1991). Assim,

$$x_j \log_2\left(\frac{x_j}{y_j}\right) = \begin{cases} 0, & \text{se } x_j = 0. \\ \infty, & \text{se } x_j \neq 0 \quad e \quad y_j = 0. \end{cases} \quad (31)$$

Da Equação (31) têm-se duas opções para o tratamento dos dados quando os vetores \mathbf{x} e \mathbf{y} apresentarem coordenadas $x_j \neq 0$ e $y_j = 0$, respectivamente. A primeira opção é assumir $x_j \log_2(x_j/0) = \gamma$, em que γ é um valor suficientemente grande. Outra maneira é considerar $y_j \rightarrow 0$, assumindo $y_j = \varepsilon$, em que ε é um valor bem próximo de zero.

Pode-se ainda garantir o domínio da função logarítmica mediante um deslocamento das coordenadas dos vetores \mathbf{x} e \mathbf{y} obtido somando um valor $\alpha \in \mathbb{R}$ em cada coordenada dos vetores, ou seja, $\mathbf{x} = (x_1 + \alpha, x_2 + \alpha, \dots, x_d + \alpha)$ e $\mathbf{y} = (y_1 + \alpha, y_2 + \alpha, \dots, y_d + \alpha)$.

Entretanto, para a aplicação da divergência KL, os vetores \mathbf{x} e \mathbf{y} devem pertencer ao conjunto de domínio d -Simplex. Assim é necessário garantir que, os deslocamento das coordenadas seja realizado de forma que a soma destas sejam iguais a 1. Considerando $\sum_{j=1}^d x_j = 1$ e $\alpha = \frac{1}{d}$, temos:

$$\begin{aligned} \sum_{j=1}^d (x_j + \frac{1}{d}) &= (\sum_{j=1}^d x_j + \sum_{j=1}^d \frac{1}{d}) \\ &= (\sum_{j=1}^d x_j + \frac{1}{d} \sum_{j=1}^d 1) \\ &= (1 + \frac{1}{d}d) = 2. \end{aligned} \quad (32)$$

Assim, sugere-se que ao aplicar a divergência KL as coordenadas dos vetores \mathbf{x} e \mathbf{y} sejam modificadas por: $\frac{1}{2}(x_j + \frac{1}{d})$ e $\frac{1}{2}(y_j + \frac{1}{d})$, respectivamente.

Generalizando este procedimento considera-se as coordenadas de \mathbf{x} , modificadas por $\frac{1}{\beta}((\beta - 1)x_j + \frac{1}{d})$ em que β é uma constante qualquer maior que 1, isto é: $\beta > 1$. Sabendo

que a $\sum_{j=1}^d x_j = 1$ tem-se que:

$$\begin{aligned}
 \sum_{j=1}^d \frac{1}{\beta} ((\beta - 1)x_j + \frac{1}{d}) &= \frac{1}{\beta} \sum_{j=1}^d ((\beta - 1)x_j + \frac{1}{d}) \\
 &= \frac{1}{\beta} \left(\sum_{j=1}^d (\beta - 1)x_j + \sum_{j=1}^d \frac{1}{d} \right) \\
 &= \frac{1}{\beta} \left((\beta - 1) \sum_{j=1}^d x_j + \frac{1}{d} \sum_{j=1}^d 1 \right) \\
 &= \frac{1}{\beta} \left((\beta - 1)1 + \frac{1}{d}d \right) \\
 &= \frac{1}{\beta} (\beta - 1 + 1) = 1.
 \end{aligned} \tag{33}$$

Para definir os tratamentos foram analisados o domínio das divergências de Bregman. Entretanto os mesmos tratamentos podem ser utilizados para garantir a aplicação das divergências Total Bregman, pois os domínio são coincidentes. Observe que, embora, a Total Bregman apresente um denominador, este sempre será diferente de zero. Desta forma assegurando a aplicação das DB tem-se, consequentemente, o domínio de aplicação da DTB garantido.

3.3.1 Avaliação dos tratamentos apresentados

Os tratamentos apresentados são avaliados em sistemas CBIR que utilizam as divergências de Bregman como medidas de similaridade. Nos sistemas desenvolvidos, utilizou-se na etapa de caracterização as técnicas BoVW e BoVW-SPM. Para a aplicação das divergências de Bregman faz-se necessário o tratamento dos dados, visto que os vetores de características, obtidos através destas abordagens, representam em suas coordenadas a frequência de uma determinada característica, podendo assim serem iguais a zero.

Tabela 5 – Resumo dos tratamentos definidos na seção 3.3 e notação.

Tratamento	Parâmetros	Notação
$x_j \log_2(x_j/0) = \gamma$	γ suficientemente grande	./TI
$y_j = \varepsilon$	ε próximo de zero	./ ε
$x_j + \alpha$ e $y_j + \alpha$	$\alpha \in \mathbb{R}$./D
$\frac{1}{\beta}((\beta - 1)x_j + \frac{1}{d})$ e $\frac{1}{\beta}((\beta - 1)y_j + \frac{1}{d})$	$\beta > 1$; d=dimensão dos dados	./N

Nesta seção temos os resultados obtidos empregando as divergências de Bregman considerando os tratamentos apresentados na Tabela 5. Além disso, são apresentados os resultados de recuperações realizadas com as distâncias Cosseno e Euclidiana.

Nos experimentos utilizou-se o banco Caltech101 com as caracterizações BoVW e BoVW-SPM, sendo utilizadas como consulta 10% das imagens de cada classe. Os parâmetros γ , ε , β , α adotados nos tratamentos são configurados considerando as características dos dados.

O desempenho das divergências empregadas na recuperação foi avaliado utilizando as médias aritméticas de todas as consultas. As métricas de avaliação empregadas foram:

Mean Average Precision (MAP), *normalized Discounted Cumulative Gain* (nDCG) e a média de precisão em k e precisão x revocação (MP@ k). A métrica nDCG foi utilizada para as primeiras 10 (nDCG@10), 20 (nDCG@20), 30 (nDCG@30) e 100 (nDCG@100) respostas da lista ranqueada. Enquanto que para a medida MP@ k , utilizamos $k=10, 20$ e 30.

3.3.1.1 Experimento I

Neste experimento avaliamos o desempenho dos sistemas CBIR utilizando a divergência GID com os tratamentos $./\varepsilon$ e $./D$, representadas respectivamente por (GID/ ε) e (GID/D). Foram considerados os parâmetros: $\varepsilon = 0,25$ e $\alpha = 1$ e para a etapa de caracterização a abordagem BoVW. A Tabela 6 apresenta uma comparação entre os resultados obtidos com esta divergência e com a distância Cosseno.

Tabela 6 – Resultados obtidos com abordagem BoVW no banco Caltech101 com as funções Cosseno, GID/D e GID/ ε .

Abordagem	BoVW			
	Função de similaridade	Cosseno	GID/D	GID/ ε
MP@10	0,2439	0,3321 +36,16%	0,3265 +33,87%	
MP@20	0,1822	0,2682 +47,20%	0,2639 +44,84%	
MP@30	0,1601	0,2423 +51,34%	0,2366 +47,78%	
nDCG@10	0,3251	0,4097 +26,02%	0,4059 +24,85%	
nDCG@20	0,2582	0,3425 +32,65%	0,3392 +31,37%	
nDCG@100	0,1593	0,2271 +42,56%	0,2212 +38,86%	
MAP	0,0890	0,1196 +34,38%	0,1170 +31,46%	

Analizando os resultados nota-se que os sistemas que utilizam as divergências (GID/D e GID/ ε) obtiveram um ganho máximo de 51% para a avaliação MP@30 quando comparados com a distância Cosseno, e um ganho mínimo de 24% na avaliação nDCG@10. De forma geral a GID/D e GID/ ε apresentaram bons resultados ao compará-los com a distância Cosseno, mostrando que estes tratamentos podem ser uma boa escolha para o uso da divergência GID.

O impacto do parâmetro ε no tratamento dos dados quando utilizamos GID pode ser visualizado na Figura 19. Comparando os resultados propiciados pela divergência GID/ ε com os obtidos com a distância Cosseno observamos que dependendo do valor de ε , a avaliação do MP@10 pode ser significativa ou não. Considerando $0,01 \leq \varepsilon \leq 0,25$ tem-se um bom resultado enquanto que este é muito inferior ao obtido com a distância Cosseno, quando $\varepsilon = 10$. Constatamos que com uma boa escolha para o valor de ε pode-se chegar a um ganho de 33% sobre a distância Cosseno.

A Figura 20 apresenta o impacto do parâmetro α na função GID/D comparando com a distância Cosseno com a aproximação BoVW, nota-se que também é necessário uma boa

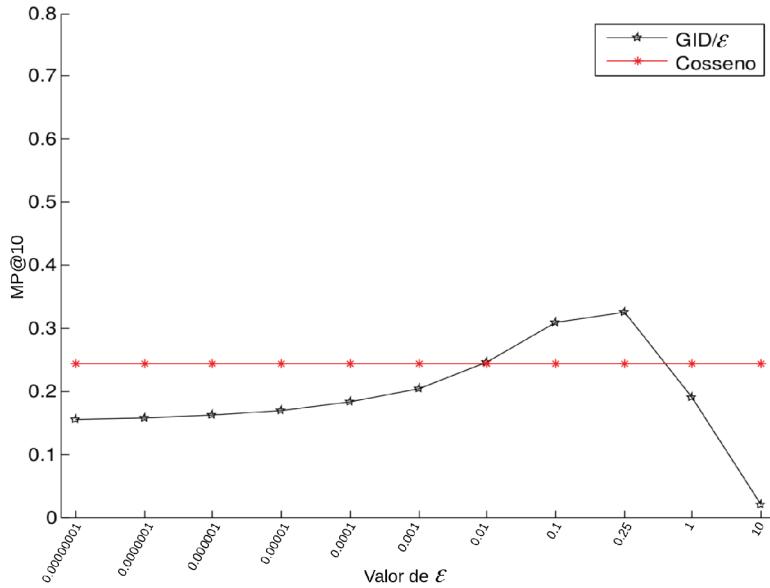


Figura 19 – O impacto do parâmetro ϵ na função GID/ ϵ , medida de avaliação MP@10 no banco de dados Caltech101 e abordagem BoVW.

escolha do α para obter ganhos de até 51% em relação a Cosseno, sendo que a partir do $\alpha \geq 1000$ o gráfico torna-se estável com a avaliação MP@10, e ao mesmo tempo superior a distância Cosseno.

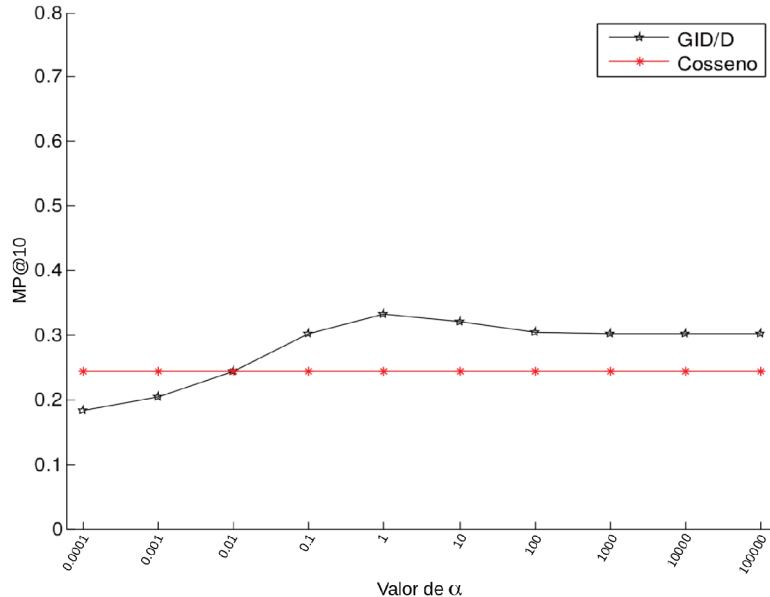


Figura 20 – O impacto do parâmetro α na função GID/D, medida de avaliação MP@10m no banco de dados Caltech101 e abordagem BoVW.

3.3.1.2 Experimento II

Neste experimento apresentamos uma comparação entre o desempenho dos sistemas CBIR utilizando os tratamentos propostos para KL e GID. A Tabela 7 apresenta os resultados e os compara com as distâncias Cosseno e Euclidiana.

Tabela 7 – Resultados obtidos no banco Caltech101 com as funções Euclidiana, Cosseno, KL e GID.

Abordagem	BoVW-SPM									
	Fun. de simil.	Eucli.	Cos.	KL/TI	KL/ ε	KL/N	GID/D			
MP@10	0,42	0,44	0,17	-61,36%	0,60	+36,36%	0,58	+31,82%	0,58	+31,82%
MP@20	0,34	0,37	0,12	-67,57%	0,52	+40,54%	0,51	+37,84%	0,50	+35,14%
MP@30	0,31	0,33	0,10	-69,70%	0,49	+48,48%	0,47	+42,42%	0,47	+42,42%
nDCG@10	0,50	0,52	0,25	-51,92%	0,65	+25,00%	0,64	+23,08%	0,64	+23,08%
nDCG@20	0,42	0,45	0,19	-57,78%	0,59	+31,11%	0,57	+26,67%	0,57	+26,67%
nDCG@30	0,38	0,41	0,16	-69,98%	0,55	+34,15%	0,54	+31,71%	0,53	+29,27%
MAP	0,12	0,15	0,10	-33,33%	0,28	+86,67%	0,27	+80,00%	0,26	+73,33%

Nos sistemas avaliados consideramos para a caracterização a abordagem BoVW-SPM e os tratamentos KL/ε , KL/N , GID/D e KL/TI , com os parâmetros: $\gamma = 10$, $\varepsilon = 0,00001$, $\beta = 3,75$, $\alpha = 1/d$, em que d é o números de bins do histograma, neste caso $d = 12.000$.

Analizando os resultados mostrados na Tabela 7, observa-se que os tratamentos propostos quando comparados com a distância Cosseno propiciaram um ganho mínimo de 23% e um ganho máximo de 86% considerando a medida de avaliação MAP. Em média, todos os tratamentos propostos (GID/D , KL/ε e KL/N), apresentaram bons resultados, excedendo os apresentados pelas distâncias Euclidiana e Cosseno, demonstrando, assim a viabilidade do uso das divergências com tratamentos e parâmetros configurados adequadamente.

A Figura 21 apresenta o impacto do parâmetro ε no tratamento dos dados quando é utilizada a divergência KL com o tratamento $./\varepsilon$ (KL/ε). Ao compararmos o desempenho obtido considerando a KL/ε com os resultados obtidos com a distância Cosseno e com a divergência KL com o tratamento $./TI$, (KL/TI) notamos que o parâmetro ε influencia na avaliação do resultado do MP@10. Observe que para $0,0000001 \leq \varepsilon \leq 0,0001$ a medida de avaliação propicia resultados superiores, possibilitando ganhos de até 36% se comparados com a distância Cosseno. Entretanto se $\varepsilon \geq 0,001$ os resultados são insatisfatórios para a avaliação MP@10.

A Figura 22 mostra a avaliação do MP@10 com a variação do valor de α na função GID/D comparando-a com a divergência KL/TI e distância Cosseno. Observe que, considerando $0,0000001 \leq \alpha \leq 0,01$ obtem-se resultados superiores aos obtidos com a distância Cosseno, e que dependendo do valor para α o desempenho da GID/D não é favorável, por exemplo $\alpha = 10$.

Ao avaliar os resultados obtidos na Tabela 7 com a KL/TI constatou-se que os resultados insatisfatórios são consequências do grande número de coordenadas nulas. Os vetores gerados pela aproximação BoVW-SPM têm em média 70% de suas coordenadas iguais a zero. Observe que as parcelas $x_i \log_2(\frac{x_i}{y_i})$ resultam em valores muito altos. Desta forma o resultado da $d_{KL/TI}(\mathbf{x}, \mathbf{y})$ não será computável.

A Figura 23 ilustra a influência dos diferentes tratamentos. Apresenta-se a recuperação de três imagens consultas contidas no banco Caltech101, em que são retornadas

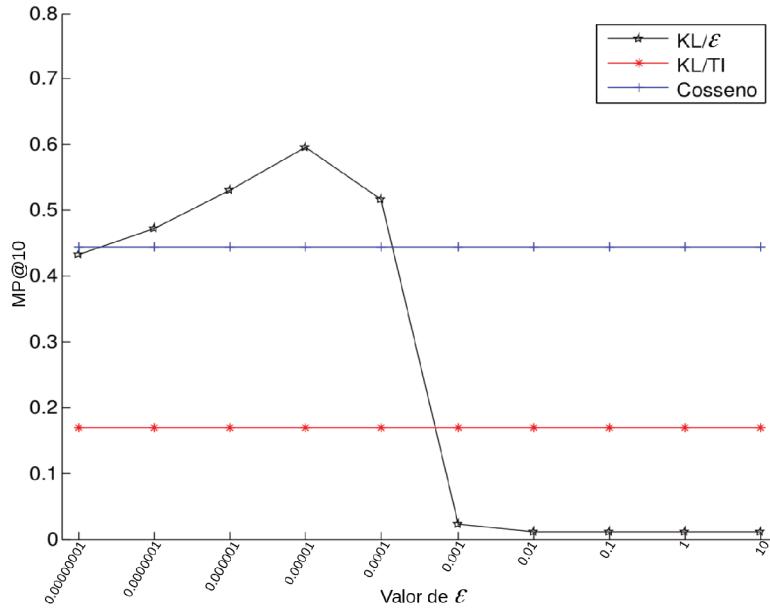


Figura 21 – O impacto do parâmetro ε na função KL/ε , medida de avaliação MP@10 no banco de dados Caltech101.

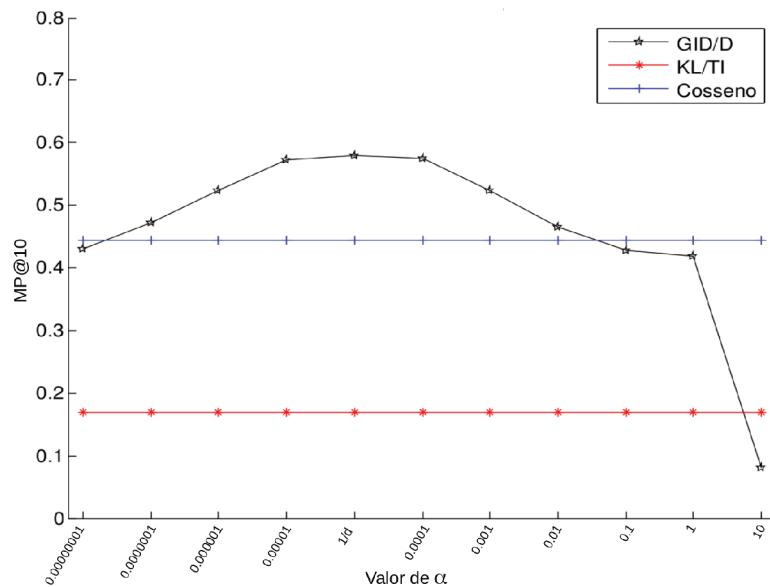


Figura 22 – O impacto do parâmetro α na função GID/D, medida de avaliação MP@10 no banco de dados Caltech101.

as 10 primeiras imagens mais semelhantes. A primeira e a segunda linha representam a recuperação KL/N e KL/TI respectivamente. Observa-se que os resultados das recuperação KL/N , linhas (a; c; e) apresentam nas 10 primeiras posições 10, 9 e 10 imagens da mesma classe que a imagem consulta, respectivamente. Na consulta utilizando KL/TI são recuperadas, respectivamente, 8, 1 e 1 imagens da mesma classe da imagem consulta.

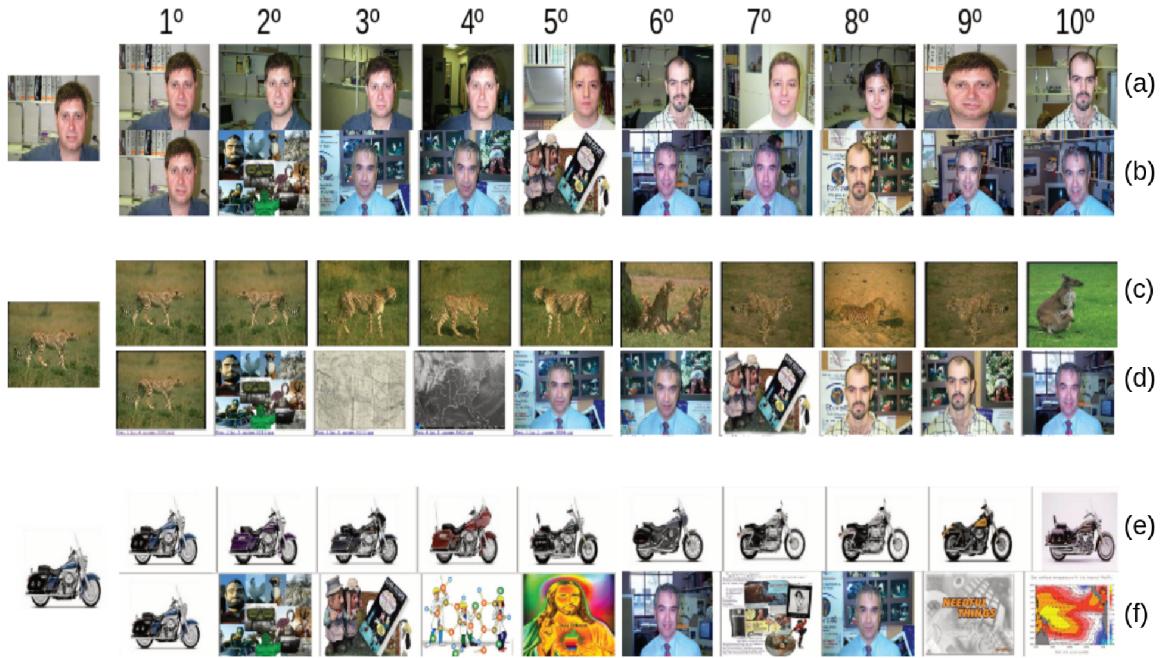


Figura 23 – Resultados da recuperação obtida aplicando os tratamentos KL/N e KL/TI

3.4 Considerações Finais

Tendo em vista o objetivo proposto neste trabalho, foram abordados, neste capítulo, as definições e propriedades das divergências de Bregman, Total Bregman.

A aplicação dos conceitos apresentados exige o tratamento adequado dos dados a fim de assegurar o domínio de aplicação destas divergências. Desta forma, novos tratamentos foram propostos e avaliados mediante experimentos empíricos.

Os resultados obtidos foram quantificados utilizando as medidas Precisão, Revocação, MAP e precisão em k e comparados com os obtidos adotando medidas de similaridades convencionais. Verificou-se que as divergências aplicadas com os tratamentos propostos apresentaram bons resultados superando os obtidos pelas distâncias convencionais. Esta análise do domínio de definição das divergências, bem como os resultados dos experimentos apresentados neste capítulo foram parcialmente apresentados em (ROCHA, 2015).

As DTB são invariantes com relação a rotação dos eixos de coordenadas e permitem a definição de centros de divergência, robustos a dados ruidosos e atípicos. Estas propriedades motivaram a definição de uma nova estrutura desenvolvida utilizando a BB-tree e a divergência Total Bregman.

CAPÍTULO 4

Busca por vizinhos utilizando árvores de bolas Total Bregman

4.1 Introdução

A busca por vizinhos próximos (NNS) é uma etapa importante em diversas aplicações tais como recuperação da informação, clusterização e classificação, dentre outras. Em espaços de alta dimensionalidade este problema tem um alto custo computacional. Desta forma, diversas estruturas têm sido desenvolvidas para armazenar uma base de dados de modo que apenas uma pequena parte desta necessite ser examinada. Os autores de (CLARKSON, 2006) e (ANDONI; INDYK, 2006) apresentam técnicas para a NNS, dentre elas destacamos os métodos baseados em estruturas de dados, as quais são utilizadas para a decomposição do espaço. Basicamente a ideia é dividir o espaço de busca em partes menores, por exemplo, retângulos, cubos, bolas ou esferas. A partição do espaço é armazenada em forma de árvore construída utilizando uma abordagem *top down*, o nó raiz corresponde à célula contendo todo o banco de dados, o qual é dividido de forma iterativa até que um critério de convergência seja satisfeito (JAIN; MURTY; FLYNN, 1999). A *kd-tree* e *Metric tree* estão dentre as principais estruturas utilizadas nesta abordagem.

A *kd-tree* utiliza a norma lp como medida de similaridade e a *Metric tree* estende a ideia básica da *kd-tree* para espaços métricos quaisquer, (CAYTON, 2008). Estas estruturas permitem uma busca rápida mediante a aplicação da estratégia *branch and bound*, (COVIELLO et al., 2013), a qual é baseada principalmente na desigualdade triangular, (JAIN; MURTY; FLYNN, 1999). Nos espaços métricos, as medidas de similaridade são estritamente simétricas, não negativas e satisfazem a desigualdade triangular. Entretanto, existem aplicações reais que envolvem, por exemplo, a comparação de probabilidades, matrizes, séries temporais, imagens, dentre outras, nas quais a noção de similaridade adequada não pode ser definida por uma métrica.

Motivado por este problema, Lawrence Cayton, em (CAYTON, 2009), apresenta uma nova estrutura com as mesmas características da *Metric tree* e um algoritmo de busca por

vizinhos mais próximos que empregam as propriedades de convexidade das divergências Bregman (DB) no lugar da desigualdade triangular. A estrutura de dados é denominada árvore de bolas de Bregman (*BB-tree*) e é obtida pela decomposição hierárquica do espaço em corpos denominados bolas de Bregman (BB). Nielsen et al em (NOCK; NIELSEN, 2009), apresentou uma extensão da *BB-tree* para divergências de Bregman simetrizadas.

Neste capítulo, propomos uma nova estrutura de dados para a NNS definida utilizando a divergência Total Bregman (DTB). A escolha desta família de divergências para a definição da nova estrutura, denominada árvore de bolas Total Bregman (*Total Bregman Ball tree - TBB-tree*), foi motivada pela sua propriedade de invariância com relação a rotação dos eixos de coordenadas e por permitir a definição de centros de divergência t-centros, robustos a dados ruidosos e atípicos, obtidos com custo e tempo computacional reduzidos, (LIU, 2011).

A DTB é obtida a partir do conceito da divergência de Bregman em (LIU, 2011) e tem sido utilizada com sucesso em diversas aplicações, como por exemplo, na recuperação da forma, (LIU et al., 2012), na análise de imagens de Tensor de Difusão (DTI), em métodos de aprendizagem de máquina, (LIU, 2011) e em técnicas de rastreamento de objetos (ROMERO; LACASSAGNE; GOUIFFES, 2013).

Para a construção da *TBB-tree* definimos corpos, a partir da divergência Total Bregman, denominados “Bolas Total Bregman”(Total Bregman Ball- TBB). A *TBB-tree* divide o espaço hierarquicamente utilizando TBBs cujos centros são robustos a dados ruidosos e atípicos. Observe que a busca na *TBB-tree*, assim como na *BB-tree*, não pode ser baseada nas propriedades de uma métrica, visto que a DTB não satisfaz a desigualdade triangular e não é simétrica. Desta forma, para realizar a busca na *TBB-tree* propomos um novo algoritmo considerando a projeção nas bolas Total Bregman. O algoritmo proposto é baseado no algoritmo apresentado em (CAYTON, 2008) o qual utiliza a DB. Na nossa proposta é utilizada a DTB que é invariante em relação à rotação dos eixos de coordenadas.

A *TBB-tree* e o algoritmo de busca propostos possibilitam a generalização dos métodos baseados em estruturas de dados para a família de divergências Total Bregman. Desta forma, permitem a seleção da divergência mais adequada para resolver um problema particular.

Neste capítulo apresentamos:

- ❑ A definição de uma nova estrutura de dados baseada nas divergências Total Bregman, a qual possibilita o desenvolvimento de meta algoritmos aplicáveis a qualquer divergência Total Bregman.
- ❑ A proposta de um novo algoritmo de busca adequado para a *TBB-tree*.
- ❑ Proposta de um algoritmo de clusterização, baseado na *TBB-tree*, válido para qualquer divergência Total Bregman. Devido às propriedades dos t-centros e das dife-

rentes escolhas de divergências, o algoritmo proposto é eficaz na clusterização de dados aglomerados em diferentes formas.

- Proposta de um algoritmo para o re-ranking de resultados obtidos utilizando sistema CBIR.

4.2 Bolas Total Bregman

Entende-se por vizinhança de p um conjunto aberto que contenha o ponto, ou seja, uma bola aberta centrada em p . Utilizando as divergências Total Bregman tem-se que uma bola aberta com raio R e centro c é dada pelo conjunto:

$$tB_\varphi(c, R) = \{x | tD_\varphi(x, c) \leq R\}.$$

Algumas bolas Total Bregman podem ser visualizadas na Figura 24. As bolas foram obtidas considerando pontos x pertencentes ao espaço \mathbb{R}^2 , $c = (0, 15, 0, 15)$ e raio $R = 0, 1$. Embora seja considerado o mesmo raio e centro, as bolas abertas têm formas e tamanhos diferentes.

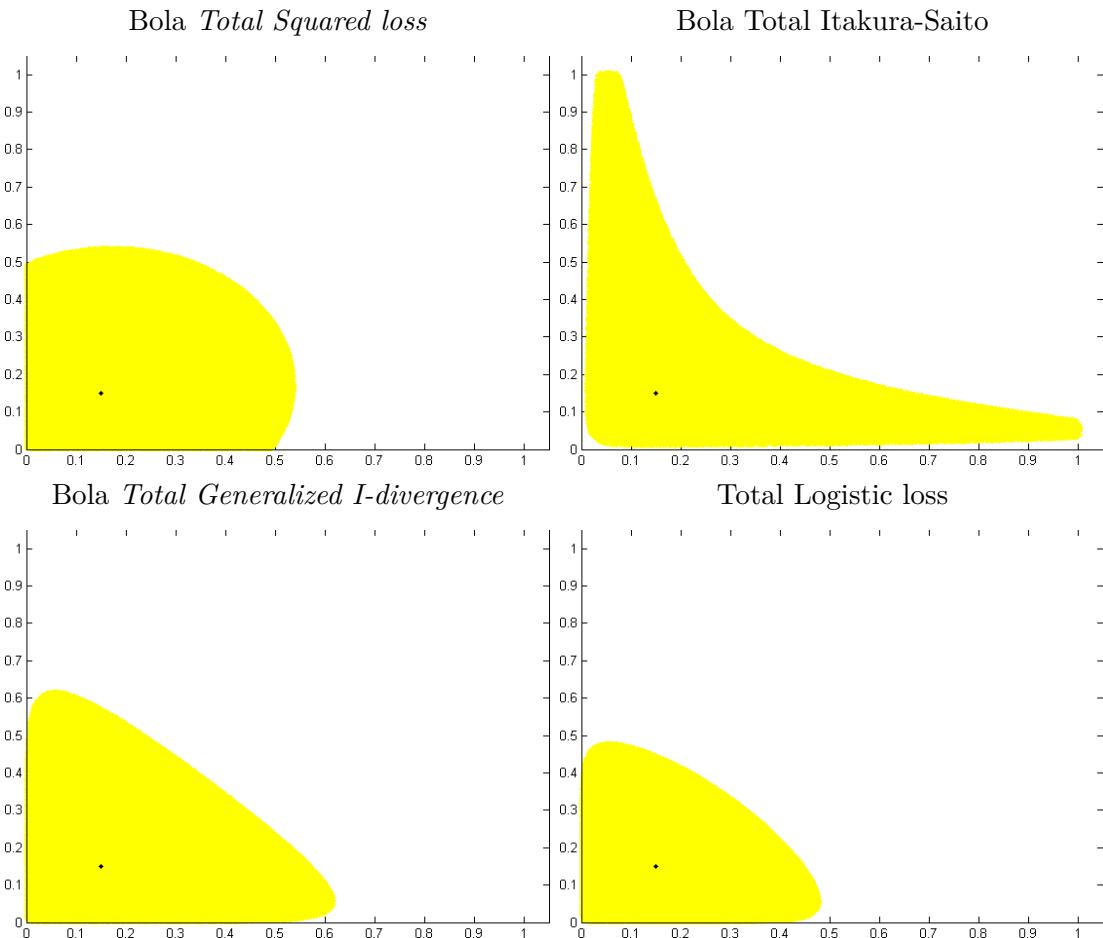


Figura 24 – Bolas Total Bregman

A partição do espaço é realizada considerando bolas cujos centros c são dados pelos pontos que minimizam a divergência Total Bregman entre eles e os demais elementos de um grupo. Conforme a seção 3.2.1, o centro, μ de um dado conjunto $X = \{x_1, x_2, \dots, x_n\}$, definido utilizando a divergência Total Bregman é denominado ***t*-centro** e dado por:

$$\mu = (\nabla \varphi)^{-1} \left(\frac{\sum w_i \nabla \varphi(x_i)}{\sum w_i} \right), \quad w_i = \frac{1}{(1 + \|\nabla \varphi(x_i)\|^2)^{\frac{1}{2}}}.$$

Observe que μ é a média ponderada do gradiente de todos os elementos de X cujos pesos w_i , são dados pelo fator conformal ρ (NIELSEN; NOCK, 2015).

4.2.1 Projeção na bola Total Bregman

Considere que $\nabla \varphi(x) = x'$. Existe uma bijeção entre x e x' dada pela transformação de Legendre definida por:

$$\varphi^*(x') = \sup_x \{ \langle x, x' \rangle - \varphi(x) \},$$

φ^* é denominada função conjugada convexa de φ . Os gradientes das funções φ e φ^* são inversamente recíprocos, ou seja, $\nabla \varphi^* = (\nabla \varphi)^{-1}$. Logo, $x = \nabla \varphi^*(x')$.

As divergências Total Bregman obtidas a partir da função geradora φ e de sua função conjugada φ^* satisfazem a seguinte propriedade: $tD_\varphi(x_1, x_2) = \rho(x_2) D_{\varphi^*}(x'_2, x'_1)$, onde $\nabla \varphi(x_1) = x'_1$ e $\nabla \varphi(x_2) = x'_2$, (NOCK; NIELSEN; AMARI, 2016).

A projeção, x_p , de um ponto, q , na bola Total Bregman, $tB(c, R)$, é o ponto da bola que minimiza a divergência $tD_\varphi(x, q)$. Desta forma deve-se resolver o seguinte problema de otimização convexa:

$$\min_x tD_\varphi(x, q) \quad \text{sujeito a } tD_\varphi(x, c) \leq R. \quad (34)$$

O Lagrangeano de (34) é

$$L(x, \lambda) = tD_\varphi(x, q) - \lambda(tD_\varphi(x, c) - R), \quad (35)$$

derivando com relação a x e igualando a zero, tem-se que

$$\nabla \varphi(x_p) = \frac{\rho_\varphi(q) \nabla \varphi(q) + \lambda \rho_\varphi(c) \nabla \varphi(c)}{\rho_\varphi(q) + \lambda \rho_\varphi(c)}.$$

Tomando,

$$\theta = \frac{\rho_\varphi(c) \lambda}{\rho_\varphi(q) + \lambda \rho_\varphi(c)}, \quad \text{ou seja, } \lambda = \frac{\rho_\varphi(q)}{\rho_\varphi(c)} \frac{\theta}{(1 - \theta)}, \quad (36)$$

segue que $x_p = \nabla \varphi^*((1 - \theta) \nabla \varphi(c) + \theta \nabla \varphi(q))$.

Portanto, de modo análogo ao apresentado em (CAYTON, 2009), determinar a projeção de q na bola $tB(c, R)$ é equivalente a encontrar um θ sujeito às seguintes restrições: $tD_\varphi(x, c) = R$, $\theta \in (0, 1]$ e $x_p = \nabla \varphi^*(\theta c' + (1 - \theta)q')$.

Observe que o segmento geodésico que liga c' a q' é o conjunto $\Gamma = \{x'/x' = (1 - \theta)\nabla\varphi(c) + \theta\nabla\varphi(q); \theta \in [0, 1]\}$. Sendo assim, o valor ótimo x_p corresponde a um $x' = \nabla\varphi(x)$ pertencente ao segmento Γ . A solução x' é obtida utilizando o algoritmo de aproximação *bisection search*, em seguida, aplica-se $\nabla\varphi^*$ para recuperar x_p .

A projeção do ponto de consulta na bola pode ser aproximada utilizando limites inferiores e superiores tais que:

$$a < \min_{x \in tB_\varphi(c, R)} tD_\varphi(x, q) < A.$$

O limite inferior a é obtido pela propriedade “weak duality”. Considerando o Lagrangeano obtido em (35) e o valor de λ dado em (36) tem-se que:

$$L(\theta) = tD_\varphi(x_\theta, q) - \frac{\rho_\varphi(q)\theta}{\rho_\varphi(c)(1 - \theta)}(tD_\varphi(x_\theta, c) - R). \quad (37)$$

Pela propriedade “weak duality”, $L(\theta) < \min_{x \in tB_\varphi(c, R)} tD_\varphi(x, q)$. O limite superior A decorre diretamente do problema primal, $\forall x_\theta$ tal que $tB_\varphi(x_\theta, c) \leq R$ segue que,

$$\min_{x \in tB_\varphi(c, R)} tD_\varphi(x, q) \leq tD_\varphi(x_\theta, q).$$

Os limites superiores e inferiores são computados e testados durante a aplicação do algoritmo de bisseção (ver seção 4.3.1).

4.3 TBB-tree

A TBB-tree é construída de forma similar à BB-tree: utiliza-se uma abordagem *top-down* e o algoritmo 2-means (*k-means* com $k=2$) para dividir hierarquicamente o espaço. Em primeiro lugar, o nó raiz é criado a partir do conjunto de dados \mathbf{X} . Em seguida o algoritmo 2-means divide \mathbf{X} em dois *clusters* utilizando os “t-centros” e a divergência Total Bregman escolhida. Este processo é repetido recursivamente até que um número máximo de pontos, previamente determinado, seja armazenado nas folhas. Cada nó interno contém duas bolas Total Bregman, uma a esquerda e outra a direita de forma que todo conjunto de pontos seja armazenado nas sub-árvores.

A árvore de Bolas Total Bregman, a partir da escolha da DTB, possibilita a partição do espaço utilizando diferentes formas na representação de cada parte. A Figura 25 ilustra a partição da região do plano, $[0, 1] \times [0, 1]$, utilizando TBB-trees com três diferentes divergências Total Bregman: *Total Squared loss*, *Total Generalized I-divergence* e *Total Itakura Saito*.

Deve-se considerar a eficiência e eficácia para estabelecer a divergência entre um elemento de consulta e estas células. A próxima seção apresenta um algoritmo de busca que estende a ideia apresentada em (CAYTON, 2009) utilizando a estrutura TBB-tree e a projeção na bola total Bregman.

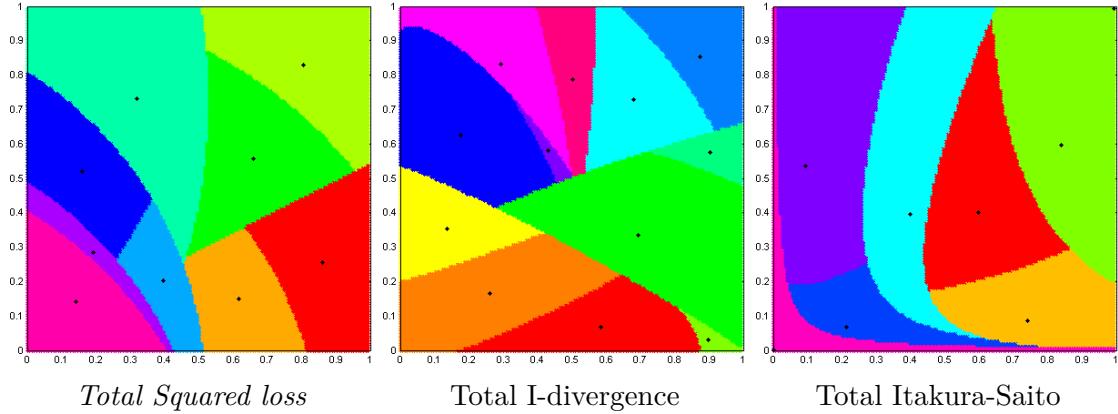


Figura 25 – Partição do Espaço utilizando TBB-trees

4.3.1 Busca com TBB-tree

Considerando a base de dados $X = \{x_1, x_2, \dots, x_n\}$, a consulta q e tD_φ uma divergência Total Bregman fixa, deve-se determinar um vizinho x_L , tal que,

$$x_L = \underset{x \in X}{\operatorname{argmin}} \quad tD_\varphi(x, q).$$

A TBB-tree é construída e percorrida em profundidade da raiz até as folhas. São escolhidos os ramos cuja “bola” tem seu centro mais próximo de q (o nó irmão é temporariamente ignorado). Uma vez determinado o nó folha mais próximo, um candidato x_c é selecionado dentre os pontos armazenados nesta folha utilizando método força-bruta. Neste momento deve-se decidir se os nós ignorados devem ser consultados ou não. O nó irmão j (ignored sibling) deve ser explorado se:

$$\min_{x \in tB_\varphi(c_j, R_j)} tD_\varphi(x, q) < tD_\varphi(x_c, q) \quad (38)$$

O lado esquerdo da desigualdade é um problema de minimização que pode ser resolvido utilizando a desigualdade triangular caso a medida de similaridade adotada seja uma métrica. Entretanto para a maioria das divergências Total Bregman esta propriedade não é válida.

Desta forma, faz-se necessário o cálculo da projeção de q na bola Total Bregman conforme apresentado na seção 4.2.1. O Algoritmo 1 apresenta o método utilizado no processo de decisão.

Algoritmo 1 CanPrune TBB-tree.

Entrada: $\theta_m = 0$, $\theta_M = 1$, x_c , q , c_j , R_j ;

início

$$\theta = \frac{\theta_m + \theta_M}{2};$$

$$x_\theta = (\nabla \varphi^*((1 - \theta)\nabla \varphi(c_j) + \theta \nabla \varphi(q));$$

$$\text{Calcule } L(\theta) = tD_\varphi(x_\theta, q) - \frac{\rho_\varphi(q)\theta}{\rho_\varphi(c_j)(1-\theta)}(tD_\varphi(x_\theta, c_j) - R_j);$$

se $L(\theta) > tD_\varphi(x_c, q)$ **então**

| Retorne Yes (não explorar o nó);

senão

se $x_\theta \in tB_\varphi(c_j, R_j)$ e $tD_\varphi(x_\theta, q) < tD_\varphi(x_c, q)$ **então**

| Retorne No (nó deve ser explorado);

senão

se $x_\theta \notin tB_\varphi(c_j, R_j)$ **então**

| Retorne CanPrune($\theta, \theta_M, q, x_c, c_j, R_j$);

senão

se $x_\theta \in tB_\varphi(c_j, R_j)$ **então**

| Retorne CanPrune($\theta_m, \theta, q, x_c, c_j, R_j$);

fim

fim

fim

fim

fim

A diferença entre o Algoritmo 1 e o apresentado em (CAYTON, 2008) é a projeção nas Bolas Total Bregman cujos centros são robustos à outliers.

4.4 Aplicações

Nesta seção a TBB-tree é utilizada para melhorar o desempenho do agrupamento de dados complexos.

4.4.1 Clustering

Os resultados de diversas pesquisas indicam que a clusterização é na verdade um método dependente da topologia dos dados. Neste contexto, torna-se importante a generalização de métodos tradicionais de forma que sejam eficientes para uma família de divergências possibilitando a seleção da melhor opção de divergência para a clusterização de um determinado conjunto de dados. Muitos algoritmos de clusterização na literatura, como *k-means*, utilizam a distância euclidiana como medida de similaridade, independente da topologia dos dados. Entretanto, esta divergência não é adequada para a clusterização de dados complexos, aglomerados em formas sobrepostas e arbitrárias.

Apresentamos um método de clusterização que utiliza a TBB-tree possibilitando a classificação dos dados utilizando as divergências Total Bregman. Devido às propriedades dos

t-centros e as diferentes formas de particionar o espaço o método permite a clusterização de dados aglomerados em diferentes formas.

A TBB-tree é utilizada para determinar uma partição do conjunto de dados, considerando a definição dos t-centros e a respectiva divergência Total Bregman. Para a construção da TBB-tree são necessários definir a divergência Total Bregman e o número máximo m de dados em cada folha da árvore denominado bucketsize.

Em seguida, agrupamentos adjacentes mais próximos são unidos até que o número de *clusters* desejado seja alcançado. A distância entre dois agrupamentos é dada pela menor distância entre seus pontos. O Algoritmo 2 resume o método proposto.

Algoritmo 2 *Clustering using TBB-tree.*

Entrada: Número de *cluster* K, bucketsize m, conjunto de dados X, opção da divergência Total Bregman D;

Saída: K grupos;

início

Construir *clusters* $C_1 \dots C_N$ utilizando a TBB-tree para X, m e D;

repita

Calcular a distância d_{ij} para cada par $C_{ij} = (C_i; C_j)$ de *clusters* adjacentes;

Selecionar o par C_{ij} tal que d_{ij} seja mínima;

Substitua C_i e C_j pela união destas regiões;

$N = N - 1$;

até $N < k$;

fim

4.4.1.1 Resultados Experimentais

O método proposto foi utilizado para classificar conjuntos de dados sintéticos complexos. Os resultados obtidos com a TBB-tree são comparados aos obtidos empregando na determinação da partição inicial a *kd-tree* e o método *k-means*, ambos utilizando a distância Euclidiana como medida de similaridade.

A TBB-tree divide o espaço em regiões com aproximadamente o mesmo número de pontos o qual é previamente determinado (*bucketsize*). Esta propriedade permite inferir que regiões com alta densidade têm menor área. Desta forma, valores altos do *bucketsize* podem definir regiões com grande área e baixa densidade, com pontos muito distantes prejudicando assim a clusterização. Nos experimentos apresentados optamos por valores baixos para definição desse parâmetro considerando-o igual a 3(três) a fim de diminuir a influência de dados ruidosos.

Na Figura 26(a) tem-se um conjunto de dados consistindo de dois arcos (dois *clusters*). Os agrupamentos obtidos utilizando o método *k-means* e a estrutura *kd-tree* são apresentados nas Figuras 26 (b) e (c), respectivamente. Observe que os dados não foram agrupados de forma correta. As Figuras 26(d), (e) e (f), ilustram os resultados da clusterização utilizando a TBB-tree com as divergências *Total Squared loss*, *Total Generalized*

I -divergence e Total Itakura Saito, respectivamente. O conjunto de dados é agrupado corretamente em dois *clusters* utilizando a divergência Total Itakura Saito.

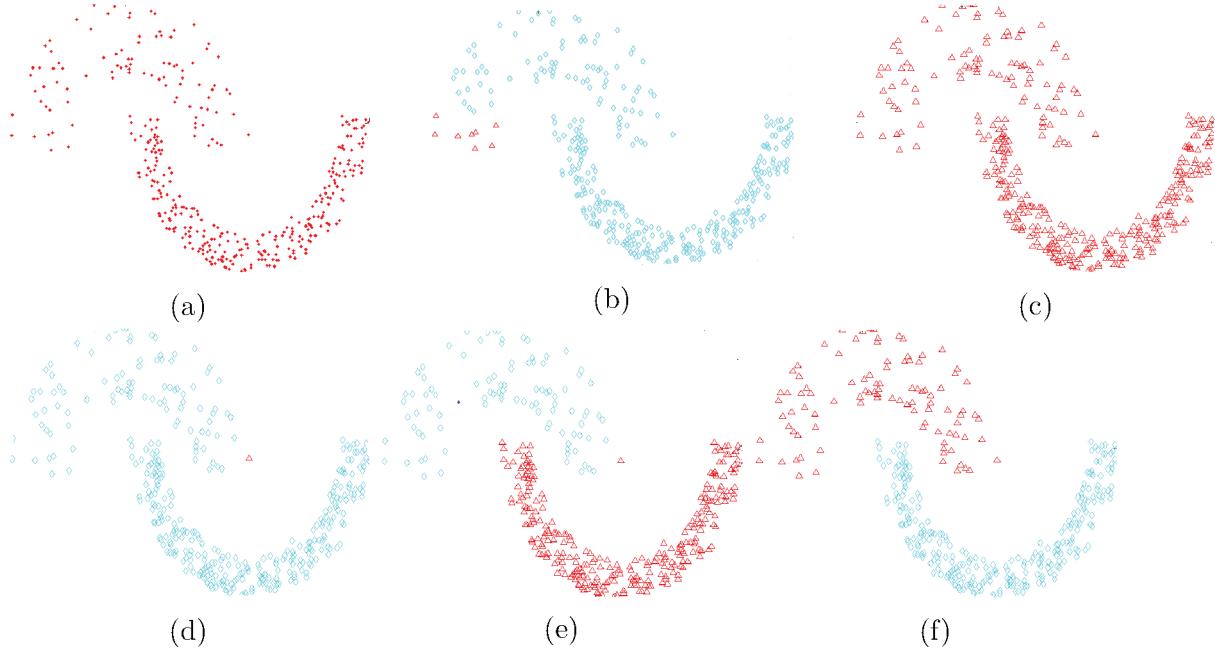


Figura 26 – (a)Conjunto de dados (JAIN; MARTIN, 2005), (b, c) *Clusters* obtidos utilizando a *kd-tree* e *k-means*, respectivamente, (d, e, f) *Clusters* resultantes utilizando a *TBB-tree* com a divergência *Total Squared loss*, *Total Itakura Saito* e *Total Generalized I-divergence*, respectivamente.

Os conjuntos iniciais obtidos com a *TBB-tree* considerando as divergências *Total Squared loss*, *Total Generalized I-divergence* e *Total Itakura Saito* podem ser visualizadas na Figura 27. Os t-centros são representados com o símbolo “o”. Observe que os t-centros representam regiões pequenas e densas e têm o formato dos agrupamentos do conjunto.

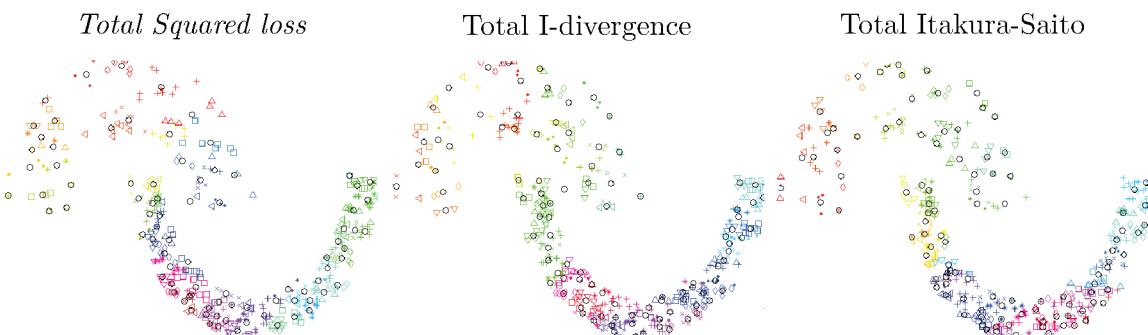


Figura 27 – Partição inicial do espaço utilizando TBB-trees e t-centros

Na próxima aplicação, dada na Figura 28, apresentamos conjuntos de dados sintéticos definidos em 2 dimensões, que contêm grupos irregulares com formas assimétricas e tamanhos diferentes. Os agrupamentos são ligeiramente sobrepostos e dificultam o desempenho dos métodos de clusterização.

Na 1^a linha o conjunto de dados utilizado tem sete *clusters*. A Figura 28(a) ilustra os agrupamentos obtidos utilizando a TBB-tree com a divergência *Total Itakura Saito* para este conjunto de dados. Os resultado obtidos podem ser comparados aos obtidos com a estrutura *kd-tree* e com o Método *k-means*, Figuras 28 (b) e (c), respectivamente.

Na 2^a linha tem-se em 28(d) os resultados da TBB-tree com *Total Squared loss* comparadas aos obtidos com 28 (e) *kd-tree* e 28(f) Método *k-means*. O conjunto de dados utilizado é formado por dois *clusters*: um círculo e um arco.

Observe que tanto a *kd-tree* quanto o *k-means* quando empregados na clusterização destes conjuntos apresenta resultados com erros de agrupamentos. Entretanto o método utilizando a TBB-tree com as divergências *Total Itakura Saito* e *Total Squared loss* são eficazes na determinação dos *clusters*.

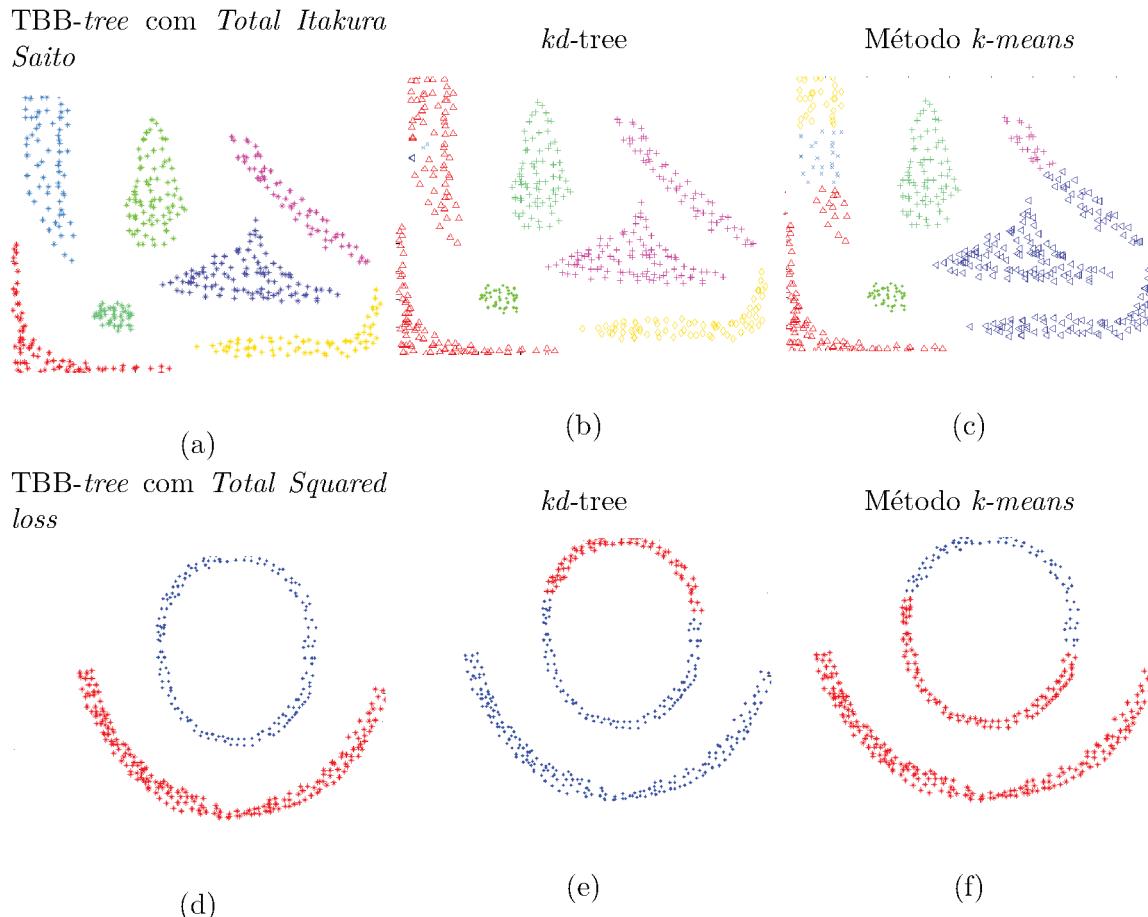


Figura 28 – Eficácia da TBB-tree na clusterização de dados complexos

4.4.2 Re-ranking

Os métodos de re-ranking têm como objetivo melhorar a eficácia do processo de recuperação. Um algoritmo de re-ranking de imagens tem como entrada um ranking inicial, utilizado para gerar um novo ranking mais eficaz.

Em (PARK; BAEK; LEE, 2005) os autores propõem um método de re-ranking que utiliza a clusterização: imagens de um ranking inicial são agrupados por meio de um método hierárquico aglomerativo e a classificação dos resultados ajustada de acordo com a distância dos grupos à imagem de consulta.

Propomos um método de re-ranking que utiliza a busca na TBB-tree para refinar o ranking inicial. O Algoritmo 3 apresenta o método proposto.

Inicialmente, dada uma imagem de consulta, obtém-se uma lista inicial de imagens (L) utilizando um sistema de recuperação. Em seguida, a TBB-tree é construída considerando as imagens do topo da lista L . Finalmente um novo ranking é gerado aplicando o algoritmo de busca apresentado na seção 4.3.1.

Algoritmo 3 Re-ranking using TBB-tree.

Entrada: Q (imagem de consulta), $L = \{L_1, \dots, L_N\}$ (Ranking inicial para Q), n (número de elementos do novo ranking), m (*bucketsize*), D (opção da divergência Total Bregman);

Saída: Ranking final;

início

```

    NumViz=0;
    Construir a TBB-tree para os n primeiros elementos da lista L, utilizando m e D;
    se  $NumViz < n$  então
        | Realizar a busca para query Q na TBB-tree conforme seção 4.3.1;
        |  $NumViz = NumViz + 1$ ;
    fim
fim

```

4.4.2.1 Resultados Experimentais

Nestes experimentos, os rankings iniciais foram gerados aplicando um sistema CBIR proposto considerando a medida de similaridade *Generalized I-divergence* e o método de caracterização *Bag of Visual Word* (BoVW). Neste sistema os histogramas h_i gerados pelas abordagem BoVW, são representados por vetores com 12000 coordenadas. Assim cada imagem é representada por um vetor de dimensão 12000 (para maiores detalhes veja (ROCHA; FERREIRA; BARCELOS, 2014)).

Apresentamos a avaliação de uma amostra dos resultados obtidos no re-ranking das imagens do banco Caltech101¹.

A Figura 29 ilustra a medida de avaliação Average Precision em $k=100$ ($AP@100$) dos resultados alcançados para 36 imagens de consulta utilizando o sistema CBIR e o método de re-ranking.

Para cada imagem de consulta, as primeiras 1000 imagens da lista gerada foram utilizadas na construção da TBB-tree com a divergência *Total Generalized I-divergence*, em seguida realizou-se a busca das 100 imagens mais próximas da query.

¹ http://www.vision.caltech.edu/Image_Datasets/Caltech101/.

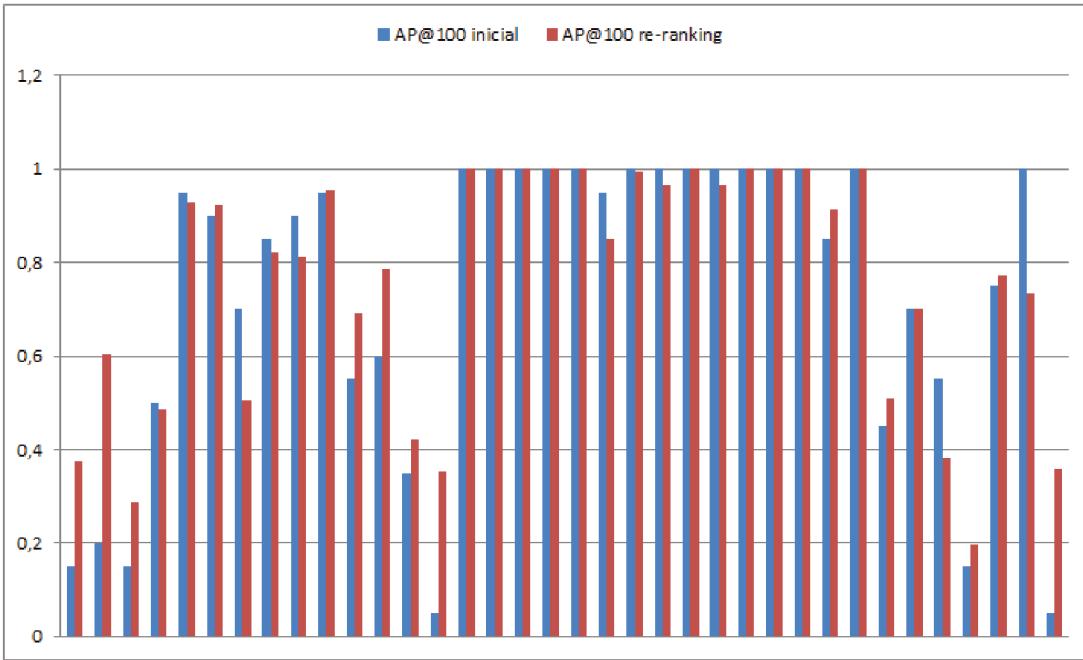


Figura 29 – Average Precision em 100.

Os resultados indicam o bom desempenho da busca realizada na árvore TBB-*tree*. A TBB-*tree* mostrou-se muito eficaz no re-ranking de imagens cuja recuperação inicial apresentava resultados com baixa precisão. Destacamos os resultados para as queries 1, 2, 14 e 36 cujo re-ranking propiciou um aumento considerável da Average Precision. Para as queries 14 e 36 a medida passa de 5%, aproximadamente, para 35%. O método foi eficaz no re-ranking de 67% da amostra. Ressalta-se que para obter este resultado consideramos apenas 1000 imagens do ranking inicial para a construção da TBB-*tree*, em torno de 10% do banco.

O re-ranking possibilitou uma variação média da medida Average Precision de 4%. O método foi aplicado uma única vez e não foi utilizada nenhuma técnica de agregação de ranking. Os resultados obtidos podem ser melhorados aplicando uma destas técnicas.

4.5 Considerações finais

Neste capítulo, apresentamos uma nova estrutura de dados e um novo algoritmo para NNS utilizando a divergência Total Bregman (DTB). A estrutura e método de busca propostos permitem a seleção do melhor candidato (DTB) para resolver um problema particular. Desta forma, a nova estrutura permite o desenvolvimento de meta-algoritmos aplicáveis a qualquer divergência Total Bregman.

Propomos dois meta-algoritmos desenvolvidos aplicando a TBB-*tree* e o método de busca no contexto da recuperação de imagens por conteúdo e clusterização. Os experimentos apresentados demonstram o bom desempenho da nova estrutura nos métodos NNS evidenciando assim a importância da nossa proposta.

CAPÍTULO 5

Divergência de Bregman Aplicada em Problemas de Segmentação Hierárquica

A segmentação de imagens é uma das primeiras etapas em qualquer processo da análise de imagens digitais e a sua precisão determina a qualidade dessa análise. O objetivo principal da segmentação é subdividir uma imagem em grupos homogêneos, denominados regiões. A homogeneidade pode ser medida em termos de cor, textura, movimento, profundidade, etc, mas para os fins deste trabalho é medido pela semelhança dos níveis de cinza. Assim as regiões são formadas por pixels, conectados ou não, agrupados mediante um critério que determina a similaridade ou dissimilaridade dos seus valores, gerando assim uma única partição.

Determinar uma segmentação para a imagem não é um problema de fácil solução, visto que o nível de detalhe em que é realizada depende do observador e do problema a ser resolvido. A análise de todas as possíveis partições de uma dada imagem é uma maneira de superar esta dificuldade. A representação de todas as possíveis regiões da imagem é denominada Partição Hierárquica \mathcal{H} da imagem. A segmentação é obtida, a partir da hierarquia \mathcal{H} construída, selecionando uma partição neste conjunto. Esta abordagem permite que sejam selecionadas regiões em diferentes escalas ou partes da imagem.

A segmentação hierárquica é geralmente construída por meio de algoritmos de “fusão” ou de “divisão e fusão”. No caso dos algoritmos de fusão, a segmentação é obtida partindo de pequenas regiões as quais são fundidas iterativamente de acordo com um critério de similaridade. Estes algoritmos são compostos por três ingredientes básicos: um modelo da região, que diz como descrever as regiões; um critério de fusão, que define quando duas regiões serão unidas ou não, e uma ordem de fusão, que diz a cada passo que algumas regiões devem ser fundidas em primeiro lugar.

Neste capítulo propõe-se a utilização do conceito da Informação de Bregman como uma alternativa para a construção da hierarquia \mathcal{H} . Demonstra-se que o método proposto é uma generalização do algoritmo de minimização da forma discreta do funcional de

Mumford-Shah (MS), diferenciando-se pelo critério de fusão e também pela medida de similaridade, definida por meio do conceito de Informação de Bregman. Desta forma, \mathcal{H} é obtida por meio de um algoritmo de fusão no qual a perda da Informação de Bregman é minimizada. A Informação de Bregman utilizada no funcional pode ser definida a partir da divergência de Bregman utilizando diferentes divergências.

5.1 Método Proposto

A ideia principal do método proposto é combinar a Informação de Bregman das regiões com as fronteiras entre elas, para obter uma partição hierárquica. Com a estrutura hierárquica construída, o problema de segmentação passa a ser localizar a partição desejada na hierarquia, o que pode ser feito estabelecendo um limiar para o número de regiões finais.

Sejam $I = \{x_i\}$ o conjunto dos pixels de uma dada imagem, $\varphi = \{R_h\}_{h=1}^k$ uma partição de I com seus respectivos representantes de Bregman, $M = \{\mu_h\}_{h=1}^k$ e K o conjunto das fronteiras entre as regiões obtidas pela partição. Para a construção da estrutura hierárquica utilizou-se um algoritmo de fusão de regiões cujo objetivo é determinar o conjunto de representantes M e fronteiras K que minimizem o funcional definido por:

$$\begin{aligned} E_\varphi(M, K) &= \sum_{h=1}^k \sum_{x_i \in X_h} v_i d_\varphi(x_i, \mu_h) + \lambda l(K). \\ E_\varphi(M, K) &= L_\varphi(\mathbf{M}) + \lambda l(K). \end{aligned} \quad (39)$$

onde:

- $E_\varphi(M, K)$ é a energia do funcional em função da imagem $X = \{x_i\}_{i=1}^n$ e das fronteiras entre as regiões K .
- $l(K)$ é comprimento das fronteiras K .
- λ é parâmetro de escala.

Observe que a segmentação pelo funcional definido na equação (39) é um problema de minimização da Perda de Informação de Bregman $L_\varphi(\mathbf{M})$, e de um termo de regularização que controla o comprimento das fronteiras K que dividem as regiões.

Devido a grande quantidade de pontos na imagem I , para obter uma partição hierárquica, construímos o Grafo de Regiões Adjacentes (RAG) dos pixels da imagem. Dessa forma, a fusão é realizada apenas para regiões vizinhas que atendam ao critério de similaridade. Para cada aresta do grafo associa-se um valor mínimo α o qual indica o custo de fusão. Depois funde-se as regiões iterativamente seguindo o critério de fusão, geralmente esse critério implica na minimização do custo de fusão. Cada vez que um par de regiões

é unido, um novo nó é criado na árvore. Este novo nó é unido ao par de nós, que agora são chamados “nós filhos”, e um novo valor α é atribuído ao “nó pai”. O algoritmo para quando existir somente uma região no RAG.

Desta forma os métodos de segmentação hierárquica se diferenciam pelas definições do custo e critério de fusão. O custo e critério de fusão utilizados neste trabalho são apresentados a seguir.

5.1.1 Custo de Fusão

A fusão de duas regiões adjacentes é interessante se o resultado desta é melhor que a representação anterior, isto é, se a energia $E_\varphi(X)$, considerando as regiões unidas, é menor do que quando, estas, estavam separadas. A partir desta afirmação o critério de fusão utilizado no algoritmo é construído.

Considere as regiões adjacentes $R_1 = \{x_1, \dots, x_l\}$ e $R_2 = \{x_{l+1}, \dots, x_m\}$ tais que $X_u = R_1 \cup R_2 = \{x_1, \dots, x_m\}$. Se \mathbf{X}_u é uma variável aleatória com valores em X_u seguindo a distribuição de probabilidade v_i o representante de Bregman de \mathbf{X}_u é $\mu = \sum_{i=1}^m v_i x_i$.

Observe que $X_d = \{R_1, R_2\}$ é uma partição de X_u e $M = \{\mu_h\}_{h=1}^2$ o conjunto dos respectivos representantes de Bregman. Se $p = \{p_h\}_{h=1}^2$ com $p_h = \sum_{x_i \in R_h} v_i$ é uma medida de probabilidade em M , tem-se a variável induzida \mathbf{M} com valores em M e distribuição de probabilidade p . As regiões \mathbf{R}_h são variáveis aleatórias com valores em R_h seguindo as distribuições de probabilidade $\frac{v_i}{p_h}$ para $x_i \in R_h$. Os representantes de Bregman das variáveis \mathbf{R}_1 e \mathbf{R}_2 são, respectivamente, dados por:

$$\mu_1 = \sum_{i=1}^l \frac{v_i}{p_1} x_i \quad e \quad \mu_2 = \sum_{i=l+1}^m \frac{v_i}{p_2} x_i.$$

O representante de Bregman, s , de \mathbf{M} , é o valor esperado de \mathbf{M} , isto é,

$$\begin{aligned} s &= \sum_{h=1}^2 p_h \mu_h = p_1 \mu_1 + p_2 \mu_2 \\ &= p_1 \sum_{i=1}^l \frac{v_i}{p_1} x_i + p_2 \sum_{i=l+1}^m \frac{v_i}{p_2} x_i \\ &= \sum_{i=1}^m v_i x_i = \mu. \end{aligned}$$

Da definição (27) tem-se que:

$$\begin{aligned} I_\varphi(\mathbf{M}) &= \sum_{h=1}^2 p_h d_\varphi(\mu_h, \mu) \\ &= p_1 d_\varphi(\mu_1, \mu) + p_2 d_\varphi(\mu_2, \mu). \end{aligned} \tag{40}$$

Observe ainda que a perda da informação de Bregman dada pela partição X_d é

$$L_\varphi(M) = I_\varphi(\mathbf{X}_u) - I_\varphi(\mathbf{M});$$

Considere as regiões R_1 e R_2 . Seja $X_u = R_1 \cup R_2$ e $M_u = \{\mu\}$ o conjunto do representante de Bregman de X_u .

Para a definição do critério de fusão calcula-se a diferença entre a energia obtida considerando a fusão das regiões, $E_\varphi(M_u, K_u)$ e a obtida quando estas permanecem separadas, $E_\varphi(M, K)$. Se o número de regiões da partição é $h=1$, ou seja não existe partição em X_u , o conjunto, K_u , das fronteiras entre as regiões é vazio. Logo $l(K_u) = 0$. Assim, da definição (5) segue:

$$\begin{aligned} E_\varphi(M_u, K_u) &= L_\varphi(\mathbf{M}_u) + \lambda l(K_u). \\ E_\varphi(M_u, K_u) &= L_\varphi(\mathbf{M}_u) = I_\varphi(\mathbf{X}_u); \end{aligned} \quad (41)$$

Seja K um conjunto unitário cujo elemento é a fronteira entre R_1 e R_2 . Daí,

$$E_\varphi(M, K) = L_\varphi(\mathbf{M}) + \lambda l(K); \quad (42)$$

$$E_\varphi(M_u, K_u) = \sum_{x_i \in R_1 \cup R_2} v_i d_\varphi(x_i, \mu) = I_\varphi(\mathbf{X}_u); \quad (43)$$

Observe que μ é o representante de Bregman da variável \mathbf{X}_u com valores em $X_u = R_1 \cup R_2 = \{x_1, \dots, x_m\}$ e medida de probabilidade v_i .

$$\begin{aligned} E_\varphi(X_d) &= \sum_{x_i \in R_1} v_i d_\varphi(x_i, \mu_1) + \sum_{x_i \in R_2} v_i d_\varphi(x_i, \mu_2) \\ &\quad + \lambda l(K); \end{aligned} \quad (44)$$

K é a fronteira entre R_1 e R_2 .

Desta forma, utilizando as equações (40),(43) e (44), define-se, como segue, o custo de fusão $\Delta_\varphi(R_1, R_2)$ das regiões R_1 e R_2 .

Definição: Sejam φ uma função relativamente diferenciável e convexa e $R_1 = \{x_1, \dots, x_l\}$ e $R_2 = \{x_{l+1}, \dots, x_m\}$ dois subconjuntos adjacente de X com representantes de Bregman em $M = \{\mu_1, \mu_2\}$, respectivamente, pertencentes ao interior de $\text{dom}(\varphi)$. Seja \mathbf{X} uma variável aleatória com valores em X seguindo a distribuição de probabilidade v_i . Sendo $p = \{p_h\}_{h=1}^2$ com $p_h = \sum_{x_i \in R_h} v_i$ uma medida de probabilidade em M e K o conjunto unitário cujo elemento é a fronteira entre as regiões. Então,

$$\begin{aligned} \Delta_\varphi(R_1, R_2) &= E_\varphi(M_u, K_u) - E_\varphi(M, K). \\ \Delta_\varphi(R_1, R_2) &= p_1 d_\varphi(\mu_1, \mu) + p_2 d_\varphi(\mu_2, \mu) - \lambda l(K). \end{aligned} \quad (45)$$

Observe que o cálculo do custo de fusão não envolve todos os elementos dos conjuntos R_1 e R_2 , somente será necessário utilizar os representantes de Bregman destas regiões.

Proposição: Seja φ uma função relativamente diferenciável e convexa e dois subconjuntos disjuntos R_1 e R_2 de X com representantes de Bregman, μ_1 e μ_2 , respectivamente, pertencentes ao interior de $\text{dom}(\varphi)$. Então,

$$\begin{aligned}\Delta_\varphi(R_1, R_2) &= \left(\sum_{x_i \in R_1} v_i \right) d_\varphi(\mu, \mu_1) + \\ &\quad + \left(\sum_{x_i \in R_2} v_i \right) d_\varphi(\mu, \mu_2) - \lambda l(K).\end{aligned}\tag{46}$$

Dem. Aplicando o Lema 3.5 de (DASGUPTA; TELGARSKY, 2012) para cada uma das regiões segue que:

$$\begin{aligned}\sum_{x_i \in R_1} v_i d_\varphi(x_i, \mu_1) - \sum_{x_i \in R_1} v_i d_\varphi(x_i, \mu) &= \\ &= \left(\sum_{x_i \in R_1} v_i \right) d_\varphi(\mu_1, \mu).\end{aligned}\tag{47}$$

$$\begin{aligned}\sum_{x_i \in R_2} v_i d_\varphi(x_i, \mu_2) - \sum_{x_i \in R_2} v_i d_\varphi(x_i, \mu) &= \\ &= \left(\sum_{x_i \in R_2} v_i \right) d_\varphi(\mu_2, \mu).\end{aligned}\tag{48}$$

Somando as equações e substituindo em 47 segue o resultado. Note que se μ_1 e μ_2 são médias de R_1 e R_2 , respectivamente e μ a média do conjunto união $R_1 \cup R_2$.

Observação: Se \mathbf{X} segue uma distribuição de probabilidade uniforme, isto é, $v_i = \frac{1}{m}$, segue que

$$\begin{aligned}p_1 &= \sum_{x_i \in R_1} v_i = \sum_{i=1}^l \frac{1}{m} = \frac{l}{m}. \\ p_2 &= \sum_{x_i \in R_2} v_i = \sum_{i=l+1}^m \frac{1}{m} = \frac{m-l}{m}.\end{aligned}$$

Daí,

$$\Delta_\varphi(R_1, R_2) = \frac{l}{m} d_\varphi(\mu_1, \mu) + \frac{m-l}{m} d_\varphi(\mu_2, \mu) - \lambda l(K).\tag{49}$$

Considerando a função convexa e diferenciável $\varphi = \|\cdot\|_2$, a divergência de Bregman é a distância *Squared loss*, isto é, $d_\varphi(x, y) = \|x - y\|_2^2$. Assim

$$\Delta_\varphi(R_1, R_2) = \frac{l}{m} \|\mu_1 - \mu\|_2^2 + \frac{m-l}{m} \|\mu_2 - \mu\|_2^2 - \lambda l(K).\tag{50}$$

Tomando:

$$\begin{aligned}l &= |R_1|. \\ m - l &= |R_2|.\end{aligned}\tag{51}$$

Note ainda que:

$$\begin{aligned}\|\mu_1 - \mu\|_2 &= \frac{|R_2| \|\mu_1 - \mu_2\|_2}{|R_1| + |R_2|} \\ \|\mu_2 - \mu\|_2 &= \frac{|R_1| \|\mu_1 - \mu_2\|_2}{|R_1| + |R_2|}.\end{aligned}\quad (52)$$

Substituindo as equações (51) e (52) em (50) obtém-se:

$$\Delta_\varphi(R_1, R_2) = \frac{|R_1||R_2|}{|R_1| + |R_2|} \|\mu_1 - \mu_2\|_2^2 - \lambda l(K); \quad (53)$$

A equação (53) é a forma discretizada do funcional Mumford-Shah, um dos métodos mais robustos utilizado na segmentação de imagens (KOEPLER; LOPEZ; MOREL, 1994).

5.1.2 Critério de Fusão

A decisão de fundir duas regiões não é tomada com base na simples comparação das duas regiões, mas considerando se a união resulta em uma melhor aproximação, ou seja, considera-se o quanto $\Delta_\varphi(R_i, R_j)$ é menor que zero.

Assim fusão de duas regiões pode ocorrer para os pares de regiões adjacentes nos quais $\Delta_\varphi(R_i, R_j) < 0$, sendo que as regiões a serem fundidas são aquelas que dentre estes pares minimiza o custo de fusão, $\Delta_\varphi(R_i, R_j)$.

Resumindo, para decidir sobre a fusão das regiões seguimos as etapas:

1. Determine o conjunto C formado por todos os pares de regiões adjacentes nos quais $\Delta_\varphi(R_i, R_j) < 0$.
2. Determine R_j , para cada R_i com $(R_i, R_j) \in C$, tal que $\Delta_\varphi(R_i, R_j)$ seja mínimo.
3. Decida pela fusão de (R_i, R_j) .

5.2 Detalhes da implementação

Para a implementação do método de segmentação proposto considere $I = \{x_i\}$ o conjunto dos pixels de uma dada imagem e $\wp = \{R_k\}$ uma partição de I. O Algoritmo 4 apresenta os passos para obter a segmentação da imagem dada. Observe que o método de segmentação pode ser aplicado utilizando diferentes escolhas para a divergência de Bregman. Estas divergências são necessárias no cálculo do custo de fusão conforme apresentado na equação (46).

Algoritmo 4 Segmentação hierárquica com DB**Entrada:** Partição inicial φ ;**Saída:** Partição final;**início** Inicialize o parâmetro de escala λ ; **repita** Calcule para cada par de região adjacente o custo de fusão $\Delta_\varphi(R_i, R_j)$ conforme definido em (46); Determine o conjunto C de todos os pares de regiões adjacentes nos quais $\Delta_\varphi(R_i, R_j) < 0$; **repita** até $\exists C_{ij} \in C$ selecionado; Seleciona o par $C_{ij} = (R_i; R_j)$ conforme o critério de fusão (seção 5.1.2); Substitua R_i e R_j pela união destas regiões; Remova de C todos os pares que incluem R_i ou R_j ; Incremente λ ; até $|\varphi| > 1$;**fim**

A partição inicial pode ser obtida considerando cada pixel da imagem uma região, ou seja, fazendo $R_i = x_i, \forall i = 1...mn$. O valor de λ depende da escolha divergência de Bregman e seu incremento ser linear, polinomial ou exponencial. O algoritmo efetua uma segmentação multi escala, em função deste parâmetro. O incremento do valor de λ possibilita as fusões de regiões que com valores inferiores deste parâmetro permaneciam separadas.

5.3 Experimentos e Análise dos Resultados

Nos experimentos realizados observou-se a habilidade do método descrito em capturar as regiões utilizando diferentes escolhas da divergência de Bregman. Analisou-se o desempenho do método para três definições de Informação de Bregman dadas pelas divergências: *Generalized I-divergence*, *Itakura Saito* e a *Squared loss*, esta última, conforme apresentado, gera a forma discretizada do funcional Mumford-Shah.

O parâmetro λ bem como a forma com que este é incrementado depende da escolha da divergência de Bregman. Neste trabalho para as divergências *Generalized I-divergence* e *Itakura Saito* o parâmetro foi incrementado linearmente enquanto que para a divergência *Squared loss* optou-se por um incremento exponencial dado por $\lambda = 2^n$ onde n é o número de vezes que o primeiro laço do Algoritmo 4 é executado.

A Figura 30 apresenta os resultados obtidos utilizando as divergências de Bregman citadas anteriormente. Observa-se que o algoritmo teve um bom desempenho em todos modelos. Os métodos que utilizam as divergências *Generalized I-divergence* e *Itakura Saito* têm resultados semelhantes ao modelo discretizado do MS. A segmentação obtida

ao utilizar o método com a distância *Generalized I-divergence* destaca-se pela capacidade de detectar os detalhes das imagens com um tempo inferior aos outros modelos, embora apresente resultados super segmentados.

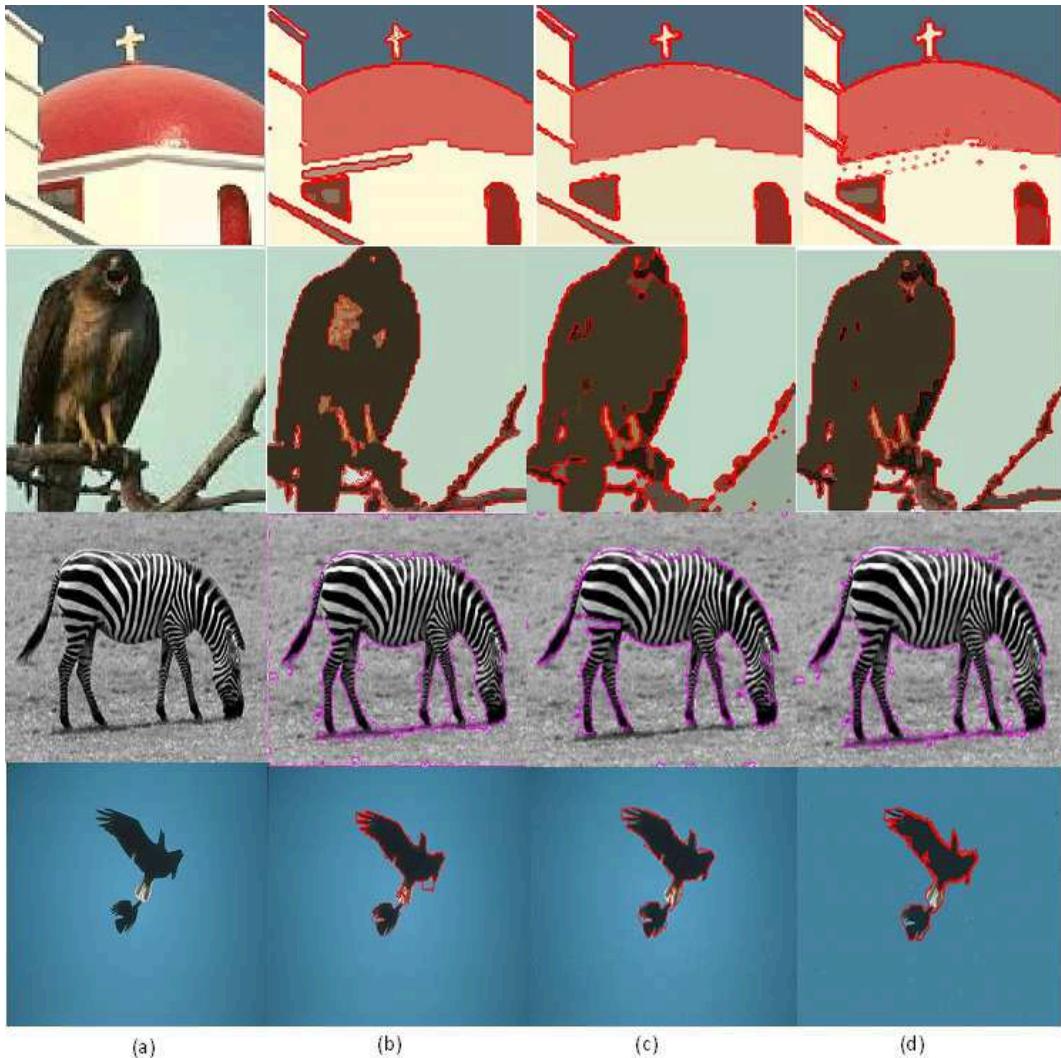


Figura 30 – (a) Imagens Originais, (b) Imagens segmentadas utilizando *Squared loss*, (c)Imagens segmentadas utilizando *Generalized I-divergence*, (d) Imagens segmentadas utilizando *Itakura Saito*.

5.4 Considerações Finais

Foi proposto a utilização do conceito da Informação de Bregman como uma alternativa para a construção de um método de segmentação hierárquica. Nesse sentido a hierarquia H é obtida por meio de um algoritmo de fusão no qual a Perda da Informação de Bregman é minimizada. A partir deste conceito o método possibilita diferentes escolhas da medida de similaridade. Se distância *Squared loss* for utilizada na definição da Informação de Bregman, o funcional apresentado é uma generalização do funcional de Mumford-Shah.

Demonstrou-se que a forma discretizada do funcional de Mumford-Shah é um caso especial do custo de fusão utilizado pelo de método segmentação hierárquica proposto.

Os resultados apresentados, obtidos utilizando as divergências *Generalized I-divergence*, *Itakura Saito* e a *Squared loss*, demonstram o bom desempenho do algoritmo para estas diferentes escolhas. A abordagem apresentada sofre com a localidade devido ao fato de utilizar um critério fusão que valida os pares de regiões um de cada vez. Isto pode ser resolvido através da validação de partições completas.

CAPÍTULO 6

Divergências de Bregman aplicadas no registro não rígido de imagens

Nas últimas décadas, uma grande quantidade de pesquisadores têm propostos modelos e técnicas computacionais para o registro de imagens devido a sua importância para diversos campos, dentre eles, análise de imagem médicas, visão computacional e sensoriamento remoto (BROWN, 1992). Os métodos de registro possibilitam a identificação de informações importantes da imagem. Por exemplo, o registro das imagens de ressonância e raio-x possibilita investigações e visualizações que podem auxiliar na detecção, diagnóstico, prognóstico, tratamento e acompanhamento de doenças (SOTIRAS; DAVATZIKOS; PARAGIOS, 2013). O registro permite ainda, acompanhar o movimento de um objeto no decorrer de uma sequência de imagens para mensurar o quanto se deslocou no decorrer da sequência (ROOS et al., 2002), (ZHANG; JIANG; TSUI, 2006).

Vários algoritmos têm sido propostos para o registro de imagens, (FITZPATRICK; HILL; JR, 2000), (GLASBEY; MARDIA, 1998), incluindo o uso de marcadores, contornos, campo de fluxo óptico, (ALVAREZ et al., 2007), caminhos difeomorfos (BARRON; FLEET; BEAUCHEMIN, 1994), aproximações estocásticas (MYRONENKO; SONG, 2010) e métodos variacionais. Entretanto a utilização dos métodos variacionais tem sido bastante relevante. Em (ALVAREZ et al., 2007) e (CAHILL; NOBLE; HAWKES, 2010) os autores demonstraram a eficácia desta abordagem.

O problema de registo de imagem é apresentado como um problema de minimização de um funcional de energia que compreende uma medida de similaridade e um termo regularizador. Diferentes técnicas de regularização tem sido inspiradas em modelos físicos, tais como modelos de corpos elásticos (DAVATZIKOS, 1997),(HE; HAMZA; KRIM, 2003), modelos de fluxo de fluido (CHIANG et al., 2008) e modelos de difusão e curvatura, (THIRION, 1998), (FISCHER; WEBER, 1993). Neste trabalho optou-se por utilizar um termo de suavização baseado na difusão.

Na literatura diversas pesquisas tem apresentado modelos variacionais com diferentes medidas de similaridade. O registo de imagens via minimização de funcional de energia

têm sido realizado utilizando o quadrado da distância *Squared loss*, (FISCHER; MODERSITZKI, 2004),(COSTA et al., 2014), a divergência Kullback-Leibler (GUETTER et al., 2005), o conceito de informação mútua, (PLUIM et al., 2003),(PLUIM; MAINTZ; VIERGEVER, 2000), (MAES et al., 2015), dentre outras.

É fato que vários algoritmos e métodos que são construídos considerando as propriedades específicas de uma determinada norma ou divergência, podem não ser diretamente aplicáveis utilizando outras funções de similaridade. Desta forma, diversas pesquisas têm sido realizadas no intuito de desenvolver métodos que sejam eficientes para uma família de funções de distância que partilham de propriedades semelhantes, (BREGMAN, 1967), (MARTIN; DURRANI, 2007). Em (PLUIM; MAINTZ; VIERGEVER, 2001), os autores apresentaram um método de registro generalizado utilizando a medida *f-information* e posteriormente o estenderam para uma série de outras medidas de similaridade, incluindo *V-information*, *I_α-information*, *M_α-information* e a medida Rényi (PLUIM et al., 2003).

Nesta perspectiva propomos um método variacional de registro que pode ser aplicado com diferentes divergências de Bregman. O problema de registro de imagens é formulado a partir da divergência de Bregman. Uma simples análise proporciona um funcional que sugere um algoritmo natural para resolver o problema do registro para divergências de Bregman arbitrárias.

Além de usar as divergências de Bregman como medidas de similaridade, também usamos informações espaciais para apoiar o processo de registro. Diversos pesquisadores têm buscado otimizar os métodos de registro considerando além da intensidade, informações espaciais (SOTIRAS; DAVATZIKOS; PARAGIOS, 2013). Em diversas situações quando a informação espacial é ignorada, o registo pode falhar pois a medida de similaridade não é capaz de quantificar a diferença entre as duas imagens. Uma forma de corrigir esta deficiência é aplicar as medidas de similaridade em um contexto local. Em (HERMOSILLO; CHEFD'HOTEL; FAUGERAS, 2002) e (CAHILL; NOBLE; HAWKES, 2010) os autores propõem o cálculo da similaridade local considerando funções densidade de probabilidade estimadas localmente ou ainda histogramas locais.

No método proposto registramos as imagens usando características que descrevem propriedades na vizinhança de cada pixel ou considerando regiões como em (GAN; CHUNG; LIAO, 2008). Desta forma as informações espaciais são incorporadas mediante a utilização de atributos adicionais que capturam outras características além da cor.

Portanto, o novo método é um meta-algoritmo que pode ser aplicado utilizando diferentes divergências de Bregman associadas a variadas caracterizações. Esta nova formulação unifica vários métodos de registro e possibilita determinarmos a melhor forma de construir e resolver o problema variacional de registro diante de novas aplicações.

A generalização da função de similaridade torna o método robusto a ruídos e aplicável no registro de imagens obtidas utilizando multimodalidades. Outro resultado interessante é que o método de registro apresentado por Fisher e Modersitzki em (FISCHER;

MODERSITZKI, 2004), segue como um caso particular de nossa abordagem.

As principais contribuições deste capítulo são:

- ❑ um novo método de registro variacional que usa as divergências de Bregman como funções de similaridade.
- ❑ uma medida de similaridade multi-característica que compara informações multidiimensionais entre as imagens modelo e de referência.
- ❑ uma formulação das equações fluxo-gradiente para divergências de Bregman considerando diferentes divergências e caracterizações.
- ❑ um modelo de registro que determina o campo de deformação para o registro de imagens obtidas a partir de diferentes scans ou modalidades.

O nosso método foi testado em conjuntos de imagens médicas com o objetivo de avaliar a qualidade do registro usando medidas específicas, como a intensidade de pixels, bem como medidas combinadas, como momentos. Também testamos o funcional proposto para várias escolhas de divergências. Nossos resultados mostram que o uso das divergências de Bregman como funções de similaridade melhoram o registro especialmente quando se utiliza a combinação de características.

6.1 Formulação do modelo de registro

Um método clássico de registro das imagens $S, T : \mathbb{R}^d \rightarrow \mathbb{R}^n$ é a minimização do funcional dado por:

$$F(u) = \mathfrak{D}(S^u, T) + \mathfrak{L}(u), \quad (54)$$

onde \mathfrak{D} é medida de distância, \mathfrak{L} representa o termo de suavização e S^u é a imagem de origem S transformada pelo campo de deformação $u(x) = u(x) = (u_1(x), u_2(x), \dots, u_d(x))$, (FISCHER; MODERSITZKI, 2004).

Utilizando como medida de similaridade a distância *Squared loss*, ou seja, $\mathfrak{D}(S^u, T) = \|S(x + u) - T(x)\|^2$ e para o termo de suavização \mathfrak{L} , o termo de difusão $\|\Delta u\|^2$, segue que :

$$F(u) = \|S(x + u) - T(x)\|^2 + \|\nabla u\|^2, \quad (55)$$

onde,

$$\|\nabla u\| \triangleq \left(\sum_{j=1}^2 \|\nabla u_j\|^2 \right)^{1/2} \text{ e } \|\nabla u_j\| = \left(\int_{\Omega} |\nabla u_j(x)|^2 dx \right)^{1/2}. \quad (56)$$

O funcional de energia definido na equação (55) utiliza a distância *Squared loss* para comparar as intensidades S e T, o objetivo, aqui, é apresentar um funcional de energia que represente o problema de registro de imagens a partir das divergências de Bregman.

Sendo assim propomos para o registro das imagens S e T a minimização do seguinte funcional:

$$F(u) = \|\nabla u\|^2 + \lambda d_\varphi(S(x+u), T(x)). \quad (57)$$

onde λ é uma constante e $d_\varphi(x, y)$ é a Divergência de Bregman entre x e y .

Para minimizar o funcional dado na equação (57) utilizamos métodos do Cálculo Variacional para determinar as equações de Euler-Lagrange e de fluxo associadas a elas. Em seguida, obtemos solução numérica utilizando diferenças finitas para discretizar as equações de fluxo.

Considerando o campo de deformação $u : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ tal que $u(x, y) = (u_1(x, y), u_2(x, y))$, da definição apresentada na equação (56) segue que:

$$\|\nabla u\| = \sqrt{\left(\frac{\partial u_1}{\partial x}\right)^2 + \left(\frac{\partial u_1}{\partial y}\right)^2 + \left(\frac{\partial u_2}{\partial x}\right)^2 + \left(\frac{\partial u_2}{\partial y}\right)^2};$$

Substituindo em (57) temos:

$$F(u) = \left(\frac{\partial u_1}{\partial x}\right)^2 + \left(\frac{\partial u_1}{\partial y}\right)^2 + \left(\frac{\partial u_2}{\partial x}\right)^2 + \left(\frac{\partial u_2}{\partial y}\right)^2 + \lambda d_\varphi(S(x+u), T(x)). \quad (58)$$

O funcional é dependente das funções $u_1(x, y)$, $u_2(x, y)$, de suas derivadas parciais e das coordenadas x e y .

$$F(u) = L(u_1, u_2, u_{1x}, u_{2x}, u_{1y}, u_{2y}, x, y), \quad (59)$$

onde $u_{ix} = \left(\frac{\partial u_i}{\partial x}\right)$ e $u_{iy} = \left(\frac{\partial u_i}{\partial y}\right)$.

Aplicando as equações de Euler-Lagrange dadas em (14) temos o sistema:

$$\left(\frac{\partial F}{\partial u_1}\right) - \left(\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_{1x}}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_{1y}}\right)\right) = 0, \quad (60)$$

$$\left(\frac{\partial F}{\partial u_2}\right) - \left(\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_{2x}}\right) + \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_{2y}}\right)\right) = 0. \quad (61)$$

Considerando que:

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_{1x}}\right) = \frac{\partial}{\partial x}(2u_{1x}) = 2\frac{\partial^2 u_1}{\partial x^2};$$

$$\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_{2x}}\right) = \frac{\partial}{\partial x}(2u_{2x}) = 2\frac{\partial^2 u_2}{\partial x^2};$$

$$\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_{1y}}\right) = \frac{\partial}{\partial y}(2u_{1y}) = 2\frac{\partial^2 u_1}{\partial y^2};$$

$$\frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_{2y}}\right) = \frac{\partial}{\partial y}(2u_{2y}) = 2\frac{\partial^2 u_2}{\partial y^2};$$

Tem-se que o sistema acima tem a seguinte forma particular,

$$-\operatorname{div}(\nabla u) + \lambda\left(\frac{\partial d_\varphi(S(x+u), T(x))}{\partial u}\right) = 0,$$

ou

$$-\Delta \mathbf{u} + \lambda \left(\frac{\partial d_\varphi(S(x+u), T(x))}{\partial u} \right) = 0.$$

Observe que $\Delta \mathbf{u} = (\Delta u_1, \Delta u_2)$, onde

$$\Delta u_i = \frac{\partial^2 u_i}{\partial x^2} + \frac{\partial^2 u_i}{\partial y^2}.$$

A próxima seção apresenta o cálculo do fluxo do gradiente para a divergência de Bregman, $\frac{\partial d_\varphi}{\partial u}$. Em seguida, utilizamos diferenças finitas para discretizar as equações de fluxo e assim obter o campo de deformação.

6.1.1 Fluxo do gradiente das divergências de Bregman

Da definição de divergência de Bregman, segue que:

$$d_\varphi(S(\mathbf{x} + \mathbf{u})), T(\mathbf{x})) = \varphi(S(\mathbf{x} + \mathbf{u})) - \varphi(T(\mathbf{x})) - \langle S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x}), \nabla \varphi(T(\mathbf{x})) \rangle$$

e portanto, as derivadas de d_φ com relação a u_i , com $i=1,2$, são dadas por:

$$\frac{\partial d_\varphi}{\partial u_i} = \langle \nabla \varphi(S(\mathbf{x} + \mathbf{u})), \nabla S(\mathbf{x} + \mathbf{u}) \mathbf{e}_i \rangle - \langle \nabla S(\mathbf{x} + \mathbf{u}) \mathbf{e}_i, \nabla \varphi(T(\mathbf{x})) \rangle;$$

simplificando a equação acima tem-se:

$$\frac{\partial d_\varphi}{\partial u_i} = \langle \nabla \varphi(S(\mathbf{x} + \mathbf{u})) - \nabla \varphi(T(\mathbf{x})), \nabla S(\mathbf{x} + \mathbf{u}) \mathbf{e}_i \rangle; \quad (62)$$

As equações de Euler obtidas considerando fluxo do gradiente para as divergências de Bregman são dadas por:

$$-\Delta \mathbf{u} + \lambda [(\nabla \varphi(S(\mathbf{x} + \mathbf{u})) - \nabla \varphi(T(\mathbf{x}))) \nabla S(\mathbf{x} + \mathbf{u})] = 0, \quad (63)$$

e sua equação de fluxo,

$$\frac{\partial u}{\partial t} = \Delta \mathbf{u} - \lambda [(\nabla \varphi(S(\mathbf{x} + \mathbf{u})) - \nabla \varphi(T(\mathbf{x}))) \nabla S(\mathbf{x} + \mathbf{u})]. \quad (64)$$

As equações obtidas em (64) são computadas para qualquer divergência de Bregman, basta determinar o gradiente das funções convexas $\varphi(x)$ a partir das quais são definidas as divergências. Assim, (63) e (64) podem ser simplificadas se considerarmos as imagens $S, T : \mathbb{R}^2 \rightarrow \mathbb{R}$, ou seja, caracterizadas apenas por um atributo.

Por exemplo, se $\mathbf{x} = (x, y)$ representa um pixel(ponto), $S(\mathbf{x})$ é a intensidade(brilho) da imagem naquele ponto. Logo S e T são funções reais e φ é função de uma variável. Portanto $\nabla \varphi = \varphi'$. Desta forma a equação (62) pode reescrita da seguinte forma,

$$\frac{\partial d_\varphi}{\partial u_i} = (\varphi'(S(\mathbf{x} + \mathbf{u})) - \varphi'(T(\mathbf{x}))) \nabla S(\mathbf{x} + \mathbf{u}) \mathbf{e}_i.$$

Se, por exemplo, $\varphi = x^2$ tem-se $\varphi' = 2x$ e consequentemente a equação de fluxo será dada por:

$$\frac{\partial u}{\partial t} = \Delta \mathbf{u} - \lambda[(2(S(\mathbf{x} + \mathbf{u})) - 2(T(\mathbf{x})))\nabla S(\mathbf{x} + \mathbf{u})]$$

como λ é constante segue que,

$$\frac{\partial \mathbf{u}}{\partial t} = \Delta \mathbf{u} - \lambda[(S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x}))\nabla S(\mathbf{x} + \mathbf{u})]. \quad (65)$$

Esta equação tem sido utilizada em diversas aplicações, (FISCHER; WEBER, 1993). A Tabela 8 contém algumas equações de Euler obtidas considerando $S, T : \mathbb{R}^2 \rightarrow \mathbb{R}$ e utilizando diferentes divergências de Bregman

Tabela 8 – Funções Convexas x Equações de Euler-Lagrange.

Dom.	$\varphi(\mathbf{x})$	φ'	Equações de Euler-Lagrange
\mathbb{R}	x^2	$2x$	$-\Delta \mathbf{u} + \lambda[(S(\mathbf{x} + \mathbf{u}) - T(\mathbf{x}))\nabla S(\mathbf{x} + \mathbf{u})] = 0$
$(0, \infty)$	$-\log x$	$\frac{-1}{x}$	$-\Delta \mathbf{u} + \lambda[\frac{-1}{S(\mathbf{x} + \mathbf{u})} + \frac{1}{T(\mathbf{x})}\nabla S(\mathbf{x} + \mathbf{u})] = 0$
$(0, \infty)$	$x \log x$	$\log x + 1$	$-\Delta \mathbf{u} + \lambda[\log(\frac{S(\mathbf{x} + \mathbf{u})}{T(\mathbf{x})})\nabla S(\mathbf{x} + \mathbf{u})] = 0$
$[0, N]$	$x \log(\frac{x}{N}) + (N - x) \log(\frac{N - x}{N})$	$\log(\frac{x}{N - x})$	$-\Delta \mathbf{u} + \lambda[\log(\frac{S(\mathbf{x} + \mathbf{u})(N - T(\mathbf{x}))}{T(\mathbf{x})(N - S(\mathbf{x} + \mathbf{u}))})\nabla S(\mathbf{x} + \mathbf{u})] = 0$

6.1.2 Caracterização utilizando informações locais

Considerando propriedades estatísticas, algébricas, geométricas, ou ainda a diferenciabilidade e as informações espaciais de uma imagem ou parte dela obtem-se características essenciais para várias tarefas do processamento de imagens.

Através da caracterização pode-se estabelecer uma comparação entre conjuntos de atributos das imagens envolvidas na análise. Assim é possível definir a similaridade entre duas imagens sem conhecer a correspondência dos pixels. Em (GOSHTASBY, 2012), os autores apresentam um estudo de 122 caracterizações sujeitas a diferentes condições de luminosidade, ruído, escala e posição geométrica. Outras revisões, classificações e avaliações deste tema podem ser encontradas em (REED; DUBUF, 1993), (TUCERYAN; JAIN et al., 1993) e (ZHANG et al., 2007).

Diversos pesquisadores têm utilizado diferentes modos para representar as informações das imagens introduzindo, através do uso de atributos, informações locais, (SOTIRAS; DAVATZIKOS; PARAGIOS, 2013). Com o intuito de capturar informações locais o algoritmo HAMMER proposto por Shen e Davatzikos em (SHEN; DAVATZIKOS, 2002) utiliza um vetor de momentos geométricos invariantes calculado em cada voxel da imagem.

Desta forma diferencia-se voxels que seriam considerados os mesmos com base apenas na informação de intensidade. Pluim et al. (PLUIM et al., 2003) propõe a utilização do gradiente da imagem como uma atributo adicional. O algoritmo proposto maximiza a informação mutua e também a intensidade do gradiente, levando em consideração tanto a magnitude do gradiente quanto a orientação.

Informações locais também podem ser incorporadas explorando uma representação de frequências locais obtidas através de filtros de Gabor, (LIU; VEMURI; MARROQUIN, 2002), (OU et al., 2011).

O modelo variacional aqui proposto permite que sejam incorporadas simultaneamente diferentes informações locais ao processo de registro. Desta forma, torna-se possível agregar informações espaciais ao método variacional de registro não rígido. Além das medidas de intensidades de cinzas, podem ser utilizadas média, entropia, momentos estatísticos, gradiente, entre outros, calculados em cada pixel considerando uma determinada vizinhança.

A construção das informações locais em um ponto x_0 pode ser realizada considerando uma vizinhança obtida utilizando largura e direção variáveis. Entretanto neste trabalho as medidas locais foram computadas a partir de um número fixo de pixels vizinhos do ponto x_0 .

Para cada característica i tem-se uma matriz S_i , cujos elementos, $S_i(x)$, são os valores de intensidade desta característica em cada ponto x da imagem dada S . Considerando uma caracterização com n informações locais as imagens de referência S e modelo T são tais que $S, T : \mathbb{R}^d \rightarrow \mathbb{R}^n$. Assim, se $x = (x_1, x_2, \dots, x_d)$, tem-se que $S(x) = (S_1(x), S_2(x), \dots, S_n(x))$ onde S_1, S_2, \dots, S_n são matrizes. De forma análoga teremos T_1, T_2, \dots, T_n matrizes cujos elementos são medidas locais obtidas a partir da imagem modelo T .

Observe que o domínio das funções convexas φ devem ser subconjuntos de \mathbb{R}^n , isto é, $\varphi : \mathbb{R}^n \rightarrow R$. Desta forma as equações (63) e (64) devem ser resolvidas considerando que:

$$\nabla \varphi(S) = \left(\frac{\partial \varphi}{\partial S_1}, \frac{\partial \varphi}{\partial S_2}, \dots, \frac{\partial \varphi}{\partial S_n} \right), \quad \nabla S(x) = \begin{bmatrix} \frac{\partial S_1}{\partial x_1} & \frac{\partial S_1}{\partial x_2} & \dots & \frac{\partial S_1}{\partial x_d} \\ \frac{\partial S_2}{\partial x_1} & \frac{\partial S_2}{\partial x_2} & \dots & \frac{\partial S_2}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial S_n}{\partial x_1} & \frac{\partial S_n}{\partial x_2} & \dots & \frac{\partial S_n}{\partial x_d} \end{bmatrix} \quad (66)$$

6.2 Detalhes da implementação

Para a implementação numérica as imagens, S e T , são representadas por n matrizes de ordem $N \times M$, cujos valores são a intensidade da característica n em cada pixel e as divergências entre elas são obtidas pelos somatórios das divergências entre cada pixel, ou seja,

$$d_\varphi(S, T) = \sum_{j=1}^m \sum_{i=1}^n d_\varphi(S(i, j), T(i, j)),$$

Seja u_{ij} , o valor do campo de deformação u associado ao pixel (i, j) com $i = 1, 2, \dots, N$ e $j = 1, 2, \dots, M$. Aplicando as equações de fluxo, dadas em (64) obtém-se imagens nos tempos $t_k = k\Delta t$ com $k = 1, 2, \dots$, sendo Δt o tamanho do passo. Considerando $u(i, j, t_k) = u_{i,j}^k$, a forma discretizada da equação (64) é dada por:

$$u_{i,j}^{k+1} = u_{i,j}^k + \Delta t \left[\Delta u_{i,j}^k - \lambda \left(\nabla \varphi(\tilde{S}(x + u_{i,j}^k)) - \nabla \varphi(T(x)) \nabla \tilde{S}(x + u_{i,j}^k) \right) \right], \quad (67)$$

onde,

$$\Delta u^0 i, j = u(i, j - 1) - 2u(i, j) + u(i, j + 1) + u(i - 1, j) - 2u(i, j) + u(i + 1, j). \quad (68)$$

Temos $\nabla S(x + u)$ dada pela equação (66) tal que

$$\frac{\partial s_p(x + u)}{\partial x}|_{ij} = s_p(x + u)(i + 1, j) - s_p(x + u)(i - 1, j) \quad (69)$$

$$\frac{\partial s_p(x + u)}{\partial y}|_{ij} = s_p(x + u)(i, j + 1) - s_p(x + u)(i, j - 1), \quad (70)$$

com $p = 1, \dots, n$.

A cada iteração para atualização de u^k , é necessário a interpolação de $S(x + u^k)$ denotada por $\tilde{S}(x + u^k)$. A interpolação é realizada em todas as matrizes de características que representam a imagem de referência S , isto é, em S_1, S_2, \dots, S_n .

A forma discretizada apresentada em (67) sugere um algoritmo iterativo (veja algoritmo 5) para a solução do problema de registro. Esta abordagem iterativa foi apresentada, inicialmente em (THIRION, 1998), na qual o campo de deformação é obtido a partir de estimativas realizadas utilizando medidas de similaridade e regularização. Em conexão com o algoritmo de Thirion, Fisher e Modersitzki em (FISCHER; MODERSITZKI, 2004) fornecem um algoritmo eficiente para o registo. O resultado é dado pela solução de um sistema obtido da linearização de uma equação diferencial parcial de difusão. O algoritmo proposto por Fisher segue como um caso particular de nossa abordagem ao utilizar como termo de similaridade a distância *Squared loss*.

6.3 Experimentos e Análise de Resultados

Nesta seção apresentam-se experimentos realizados com o intuito de avaliar a eficácia do funcional proposto em (57) utilizando diferentes escolhas tanto para a divergência de Bregman quanto para a caracterização.

A análise do modelo com diferentes divergências empregadas a diferentes informações visa verificar o impacto que estes fatores causam no registro das imagens, evidenciando assim as vantagens da generalização proposta neste trabalho.

Algoritmo 5 Registro Variacional com divergências de Bregman

Entrada: Imagens S e T, λ , função convexa φ , passo Δt , número de iterações m .

Saída: Campo de deformação u , Imagem final $S(x + u^m)$.

início

Represente as imagens S e T utilizando n matrizes de ordem $N \times M$ cujos valores são a intensidade da informação n em cada pixel.

Inicialize Δu_{ij}^t utilizando a equação (68).

para $k = 1 \dots m$ **faça**

Calcule $\nabla S(x + u)$ dada em (66) utilizando (69) e (70).

Calcule u_{ij}^{k+1} utilizando a equação (67).

Atualize u , interpolando $S(x + u^{k+1})$ (a interpolação é realizada em todas as matrizes que caracterizam S, (S_1, S_2, \dots, S_n) .

fim

fim

Considerou-se experimentos com dois fatores, denominados “Caracterização” e “Divergência”. Para o fator “Divergência” foram utilizadas as divergências de Bregman: *Squared loss*, *Logistic loss*, *Generalized I-divergence* e *Itakura Saito*. A Tabela 8 apresenta as equações de Euler referentes a cada uma destas divergências empregadas na discretização do modelo proposto. As opções do fator “Caracterização” foram definidas a partir de combinações de diversas informações das imagem. Além da matriz de intensidades de cinzas, foram utilizadas matrizes com os valores da média, entropia, momentos invariantes (LIU; RIBEIRO, 2010). Os experimentos foram realizados considerando combinações dessas características tais como: “cor e média”, “cor e momentos”, “cor e entropia”, “média e entropia”, dentre outras.

A caracterização é realizada associando a cada imagem, matrizes cujos valores são a intensidade da característica em cada pixel. As combinações utilizadas nos experimentos foram realizados considerando as seguintes características:

- ❑ cor : cada pixel da imagem é representada pelo nível de cinza.
- ❑ média: a cada pixel é associado a média calculada considerando uma vizinhança de oito pontos.
- ❑ momentos invariantes: a cada pixel é associado o valor dos momentos invariantes de Hu calculados considerando oito pontos vizinhos. Os momentos de ordem $(p + q)$, são definidos por,

$$M_{p,q} = \sum \sum x^p y^q f(x, y), \quad (71)$$

onde $p, q \geq 0$ e $f(x, y)$ é a intensidade em cada pixel (x, y) .

- ❑ entropia: a cada pixel é associado a entropia calculada considerando uma vizinhança de nove pontos.

- desvio padrão: a cada pixel é associado o desvio padrão calculado considerando uma vizinhança de nove pontos
- borda: a matriz de bordas obtida pelo algoritmo de detecção *Canny*, (CANNY, 1983).

Para avaliar a qualidade do registro realizado utilizamos as medidas de desempenho: *Correlation Coefficient* (CC) e a *Root Média Square Error* (RMSE), definidas, respectivamente em (23) e (24).

É possível calcular a correlação considerando a intensidade de qualquer informação adotada para o fator de “Caracterização” das imagens. Assim, pode-se ter o coeficiente de correlação entre as imagens considerando não só a intensidade da cor, mas também os fatores entropia ou média.

Para avaliar a evolução do CC após o registro em relação ao coeficiente de correlação inicial utilizamos o coeficiente de correlação relativo dado por:

$$CC_{\text{rel}} = \frac{(CC_{\text{final}} - CC_{\text{inicial}})}{(1 - CC_{\text{inicial}})}. \quad (72)$$

onde CC_{inicial} e CC_{final} são os coeficientes antes e após o registro, respectivamente.

O modelo de registro definido pelas equações (63) e (64) foi aplicado a imagens médicas e sintéticas, utilizando diferentes escolhas tanto para a divergência de Bregman quanto para a caracterização, com o intuito de: 1) avaliar a convergência do método e a influência do parâmetro λ , 2) analisar a interação entre os fatores “Divergência” e “Caracterização” e o efeito que estes exercem, individualmente, na qualidade do registro e 3) avaliar o desempenho do método proposto, comparando-o com o registro obtido utilizando a função “imregister” disponível no *Image Processing toolbox* do *software Matlab*.

A partir dos objetivos, os experimentos foram organizados conforme apresentado na Figura 31 e apresentados considerando as seguintes ênfases: “Avaliação da convergência do método proposto”, “Análise do impacto do valor de λ ”, “Análise do efeito do fator “Divergência” no registro de imagens”, “Análise do efeito da “Caracterização” no registro de imagens” e “Análise comparativa do funcional proposto.”

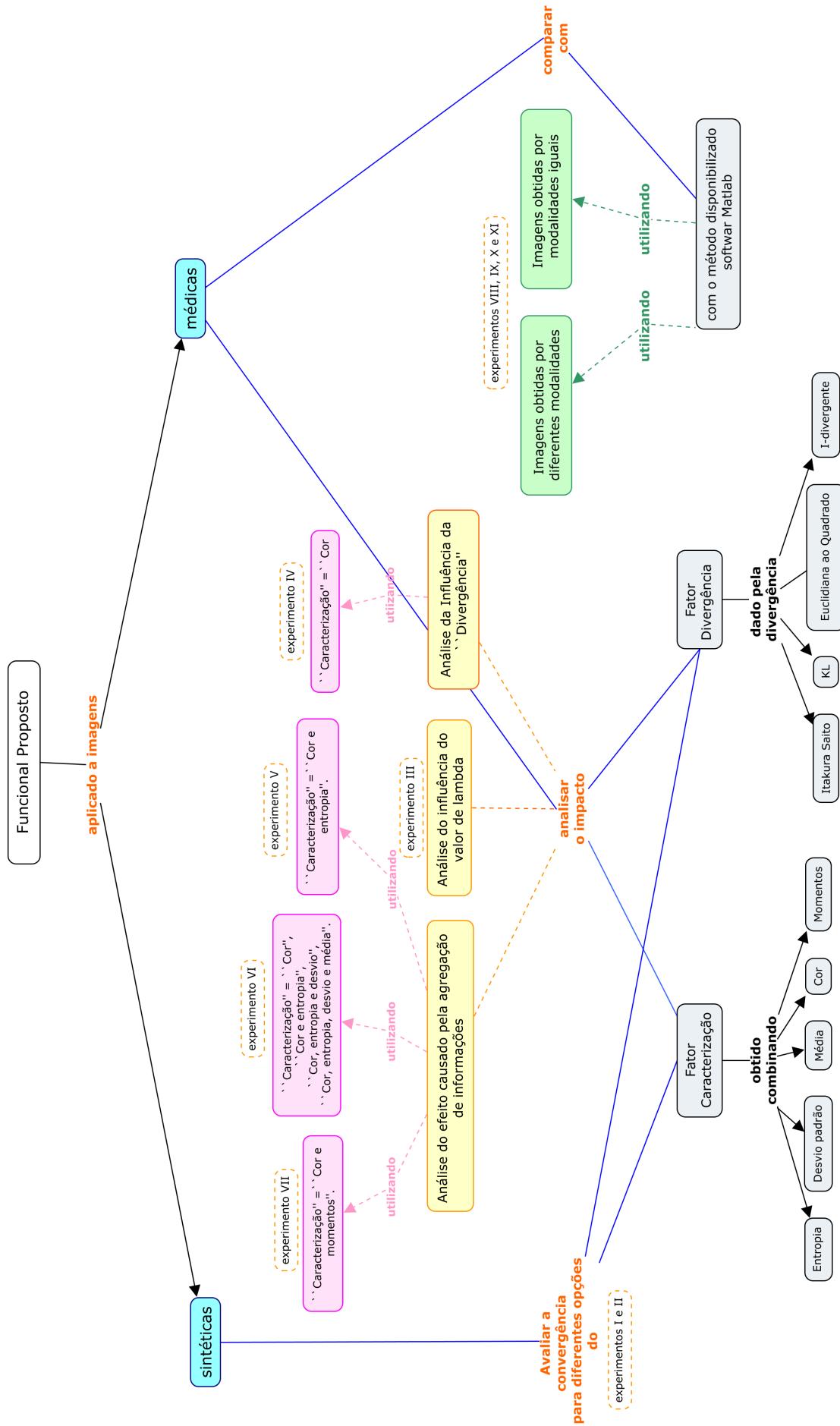


Figura 31 – Planejamento dos experimentos

6.3.1 Domínio de aplicação das divergências de Bregman

Na prática a utilização das divergências exige o tratamento dos dados, visto que é necessário, devido aos pontos de descontinuidade, garantir o domínio de aplicação das respectivas divergências. Assim, neste trabalho, sempre que necessário, será definido como domínio de aplicação o conjunto fechado $D = [0 + \epsilon, 1 - \epsilon]$. A caracterização de uma imagem é inicialmente dada por uma matriz I , cujos elementos, $I(i, j)$, são os valores de intensidade da característica em cada pixel (i, j) , tal que $I(i, j) \in [0, 1]$. No tratamento destes dados obtemos novos valores para os elementos da matriz I dados por:

$$I^{novo}(i, j) = \frac{I(i, j) - \min(I)}{\max(I) - \min(I)}(1 - 2\epsilon) + \epsilon. \quad (73)$$

O conjunto imagem das divergências dependem do domínio de aplicação definido. Por exemplo, para x, y no domínio $D = (0, 1)$ a divergência *Squared loss* varia entre 0 e 1, enquanto que o conjunto imagem das divergências Logistic loss e *Generalized I-divergence* é $(0, +\infty)$. Na prática, excluindo os pontos de descontinuidade estas divergências variam no intervalo $[0, K]$, onde K depende do valor ϵ utilizado. A Tabela 9 apresenta os intervalos de variação das divergências para diferentes valores de ϵ .

Tabela 9 – Conjunto imagem de algumas divergências de Bregman em função da escolha do ϵ

Divergências	ϵ			
	0	0.1	0.01	0.001
Logistic loss	$[0, +\infty)$	$[0, 1.76]$	$[0, 4.50]$	$[0, 6.89]$
GID	$[0, +\infty)$	$[0, 1.78]$	$[0, 3.57]$	$[0, 5.90]$
<i>Itakura Saito</i>	$[0, +\infty)$	$[0, 5.80]$	$[0, 93.40]$	$[0, 991.09]$

O comportamento apresentado na Tabela 9 pode ser visualizado na Figura 32. São apresentados os valores de divergência entre os pixels de duas imagens médicas utilizando a *Logistic loss* definida considerando $\epsilon = 0.01$ e $\epsilon = 0.001$.

6.3.2 Avaliação da convergência do método proposto para diferentes opções de “Divergência” e “Caracterização”.

Nos experimentos I e II desta seção, o modelo de registro proposto na equação 64 foi aplicado a pares de imagens sintéticas com o objetivo de observar o comportamento do método proposto mediante as diferentes escolhas realizadas para os fatores “Divergência” e “Caracterização”.

A convergência do método foi avaliada utilizando o valor do coeficiente de correlação dado em (23).

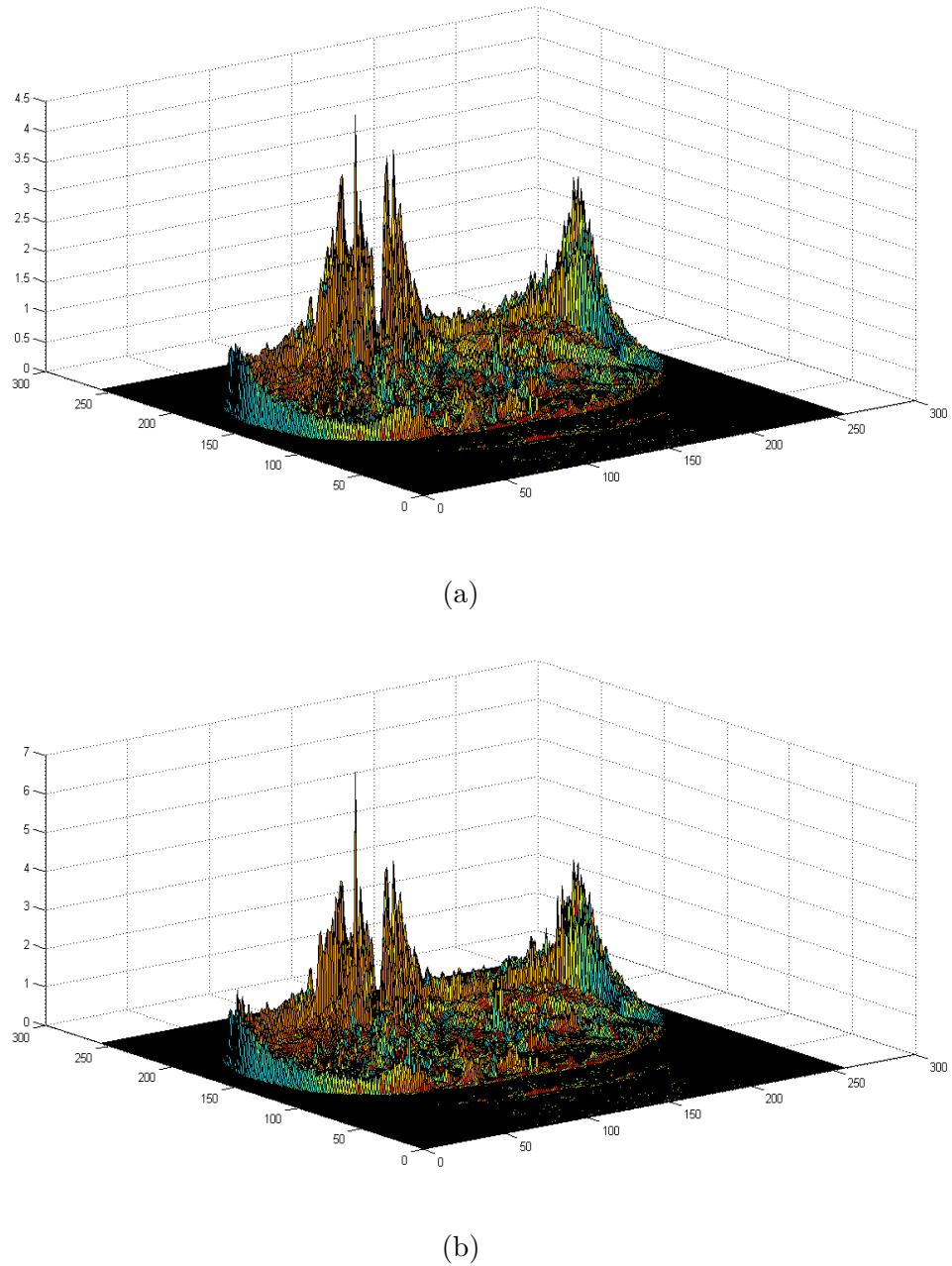


Figura 32 – BD entre os pixels de duas imagens médicas utilizando Logistic loss com (a) $\epsilon = 0.01$, (b) $\epsilon = 0.001$.

6.3.2.1 Experimento I

Neste experimento são utilizadas as imagens S e T cuja diferença inicial é apresentada na primeira linha da Figura 33. Os testes foram realizados fixando uma caracterização e variando as divergências.

Para cada combinação de característica e divergência foram realizadas 50 iterações com passo $\Delta t = 0,01$. Os valores adotados para λ considerando as divergências *Squared loss*, *Logistic loss*, *Generalized I-divergence* e *Itakura Saito*, foram 90, 10, 10 e 2, respectivamente. O coeficiente de correlação entre as imagens S e T, $CC(S, T)$, é igual a 0,8457.

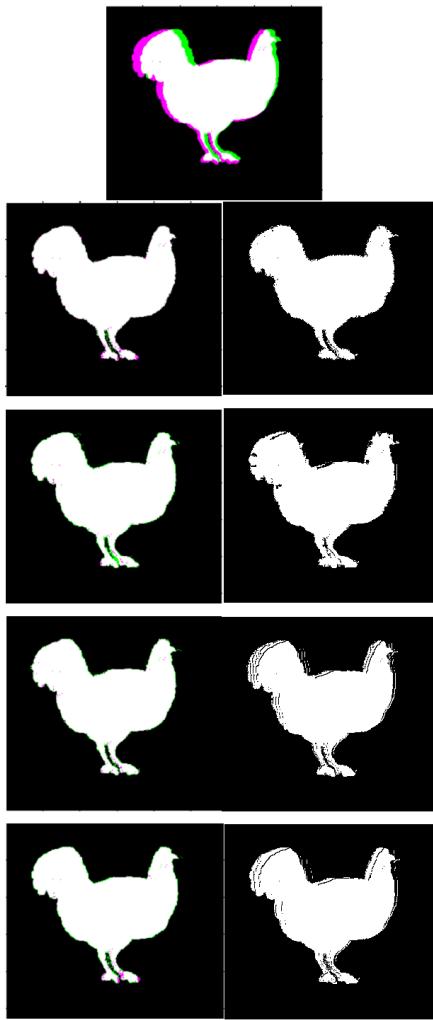


Figura 33 – Imagens Registradas x Cor. Primeira linha: Diferença inicial. Primeira coluna: Diferenças finais obtidas utilizando as divergências: *Squared loss*, *Logistic loss*, *Generalized I-divergence* e *Itakura Saito*. Segunda coluna: Imagens finais, utilizando as divergências: *Squared loss* (S^{sl}), *Logistic loss* (S^{ll}), *Generalized I-divergence* (S^{gid}) e *Itakura Saito* (S^{is}).

Caracterização “cor” associada a diferentes escolhas da Divergência

As imagens obtidas após o registro, utilizando como caracterização o nível de cinza em cada pixel e as divergências: *Squared loss*, *Logistic loss*, *Generalized I-divergence* e *Itakura Saito*, são apresentadas na segunda coluna da Figura 33.

Observa-se, pelas imagens obtidas, que o registro utilizando a matriz “cor” e a divergência *Squared loss*, neste experimento, apresentou um melhor resultado. Entretanto, quando avaliados os coeficientes de correlação entre as imagens obtidas com os outros modelos e a imagem T percebe-se que estas estão completamente relacionadas e apresentam valores bem próximos aos obtidos com a *Squared loss*.

A Figura 34 apresenta os valores destes coeficientes com relação às iterações realizadas. Os valores finais obtidos são $\text{CC}(S^{eucl}, T) = 0,9892$, $\text{CC}(S^{ll}, T) = 0,9830$, $\text{CC}(S^{gid}, T) = 0,9871$ e $\text{CC}(S^{is}, T) = 0,9847$, sendo S^{sl} , S^{ll} , S^{gid} e S^{is} as imagens finais geradas a partir da imagem de origem S e do campo de deformação $u(x)$ utilizando, respectivamente, os modelos de registro com as divergências: *Squared loss*, *Logistic loss*, *Generalized I-divergence* e *Itakura Saito*.

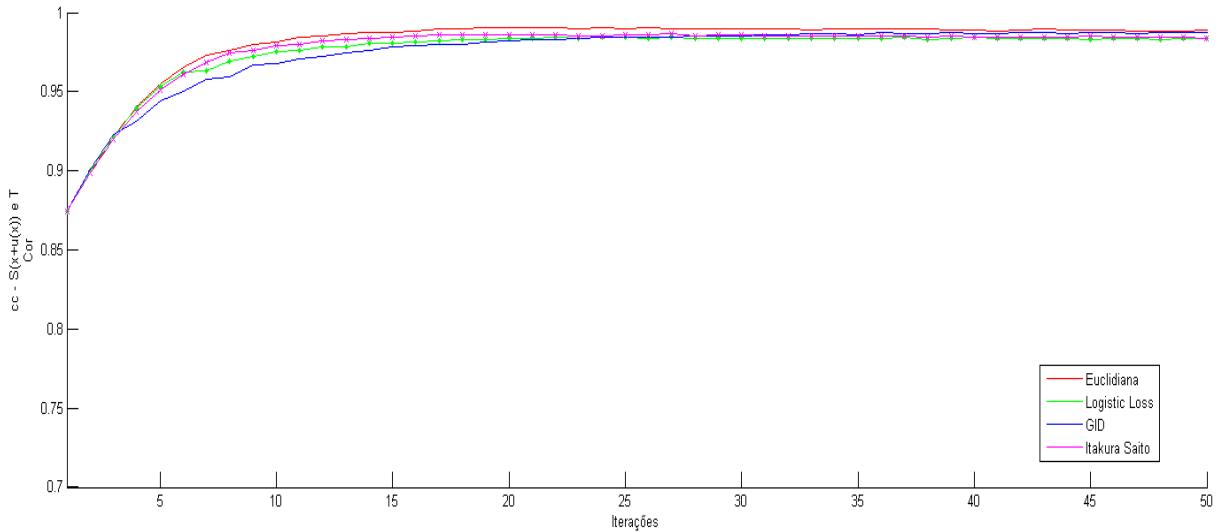


Figura 34 – Coeficiente de correlação- “cor”x Iterações.

Caracterizações “cor e entropia” e “cor e Média” associadas a diferentes escolhas da Divergência

Na primeira linha da Figura ?? são apresentadas quatro imagens obtidas utilizando para cada pixel as informações “cor e entropia” e considerando as divergências *Squared loss*, *Logistic loss*, *Generalized I-divergence* e *Itakura Saito*. De modo análogo são apresentados, na segunda linha, os resultados para cada divergência considerando a caracterização “cor e média”.

Observando as imagens apresentadas na primeira linha da Figura ??, as divergências *Itakura Saito* e *Generalized I-divergence* têm resultados bastante satisfatórios e aparentemente mais relacionados com a imagem modelo T quando comparados ao modelo que usa a distância *Squared loss*. Entretanto, analisando os valores dos coeficientes de correlação dessas imagens mostrados na Figura 36, conclui-se que, de fato, o registro realizado empregando a divergência *Itakura - Saito* apresenta melhor resultado que a *Squared loss*, enquanto que a *Generalized I-divergence* é menos relacionada com a imagem T do que esta última. É importante ressaltar que a diferença entre os coeficientes de correlação obtidos é pequena e que todas as imagens são altamente correlacionadas com a imagem modelo T, pois apresentam coeficientes de correlação superiores a 98%.

A imagem, S^{ll} , resultante do registro utilizando a divergência *Logistic loss* apresenta-se mais homogênea que as demais, entretanto um conjunto de pontos da imagem não foi

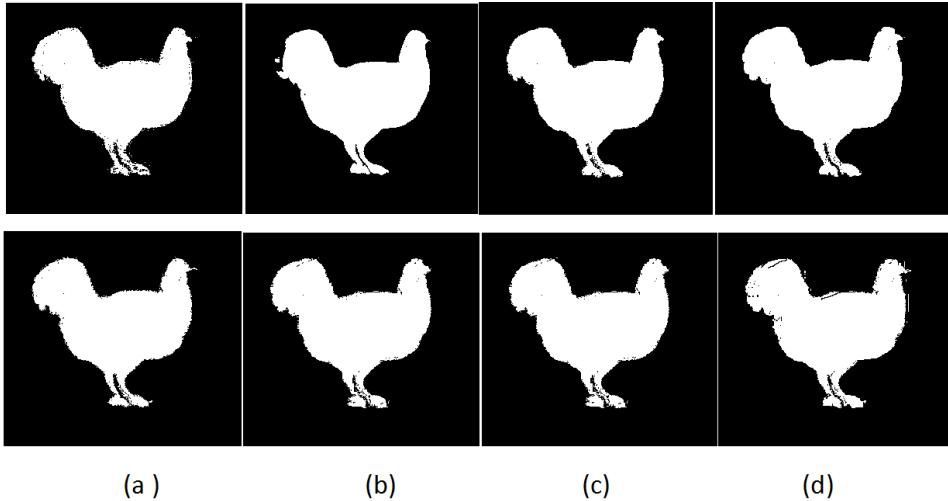


Figura 35 – Imagens Registradas com as caracterizações “cor e entropia” e “cor e Média”.

(a) *Squared loss*, (b) *Logistic loss*, (c) *Generalized I-divergence* e (d) *Itakura Saito*.

alterado, é possível que este comportamento seja consequência do número de iterações utilizado.

Observa-se que o número iterações, neste caso, foi adequado para a divergências *Squared loss* e *Itakura - Saito* já que os gráficos referentes a estas divergências apresentam um comportamento de convergência, enquanto que para as divergências *Logistic loss* e *Generalized I-divergence*, os gráficos ainda não apresentaram este tendênciam.

Para a caracterização “cor e média”, os valores finais dos coeficientes de correlação são: $CC(S^{euc}, T) = 0,9912$, $CC(S^{ll}, T) = 0,9840$, $CC(S^{gid}, T) = 0,9847$ e $CC(S^{is}, T) = 0,9867$. As divergências *Squared loss*, *Logistic loss* e *Itakura - Saito* apresentaram melhor desempenho quando comparados aos resultados obtidos com a caracterização “cor”. Os

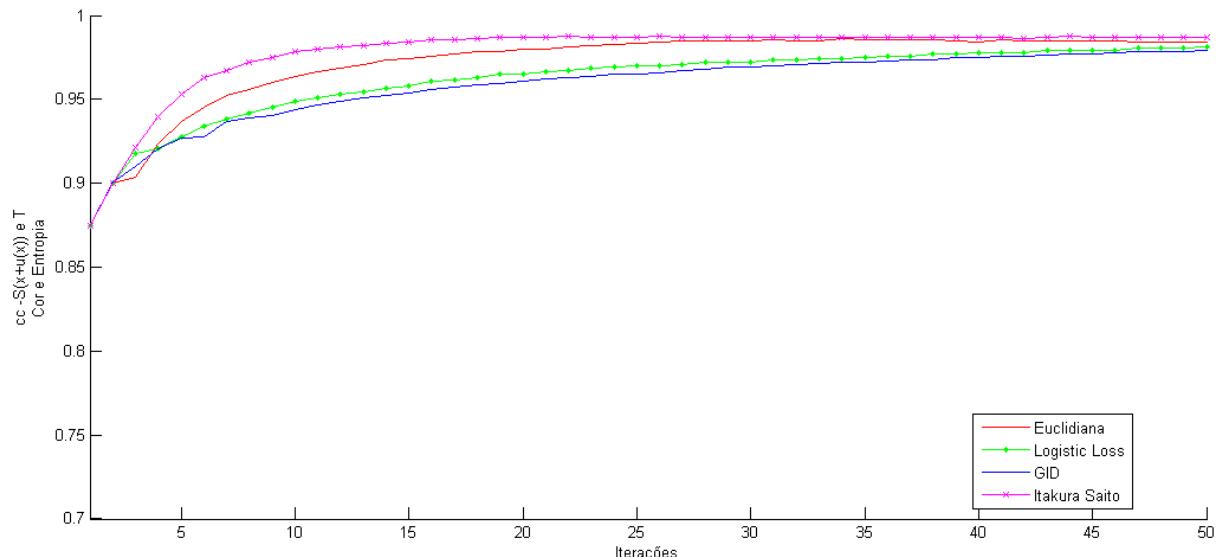


Figura 36 – Coeficiente de correlação “cor e entropia” x Iterações.

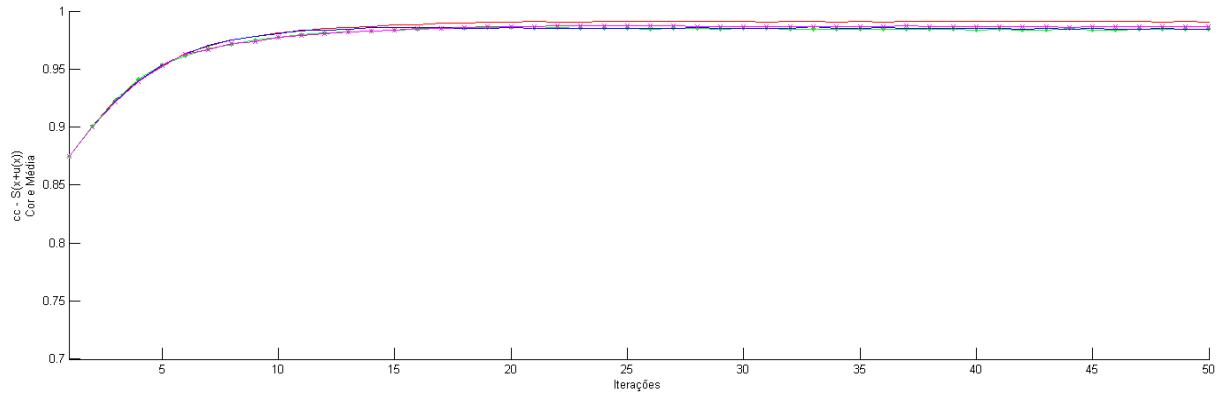


Figura 37 – Coeficiente de correlação- “cor e Média” x Iterações.

valores dos coeficientes de correlação deste teste são apresentados na Figura 37.

Caracterização “momentos” e “cor, média, momentos, entropia e borda” associada a diferentes escolhas da Divergência

Os resultados obtidos com as caracterizações “momentos” e “cor, média, momentos, entropia e borda” são ilustrados nas Figuras 38 e 39, respectivamente. Nestes testes utilizaram-se as divergências: *Squared loss*, *Logistic loss* e *Generalized I-divergence*.

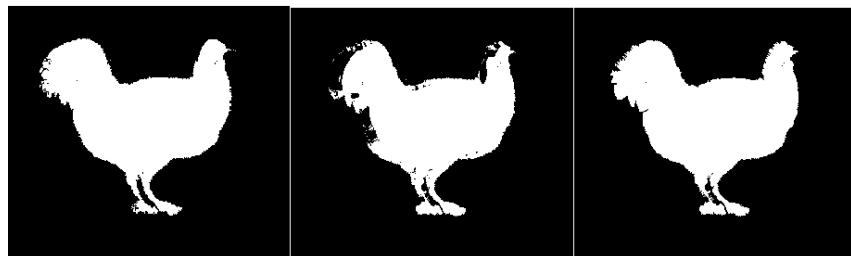


Figura 38 – Imagens Registradas com as caracterizações “momentos”. (a) *Squared loss*, (b) *Logistic loss*, (c) *Generalized I-divergence*.

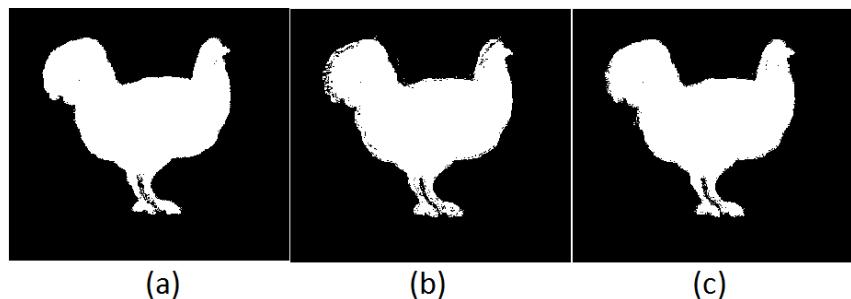


Figura 39 – Imagens Registradas com a caracterização “cor, media, momentos, entropia e borda”. (a) Distância *Squared loss*, (b) divergência *Logistic loss*, (c) *Generalized I-divergence* e (d) divergência *Itakura Saito*.

Observa-se que a divergência *Logistic loss* não apresentou um bom desempenho para as caracterizações “momentos” e “cor, media, momentos, entropia e borda”.

Análise dos resultados do Experimento I

Pode-se observar, da análise apresentada, que as divergências têm comportamentos diferentes de acordo com as caracterizações escolhidas. A divergência *Itakura Saito* obteve desempenhos bem próximos para as caracterizações “cor e entropia”, “cor e média” e “cor”, Figura 40.

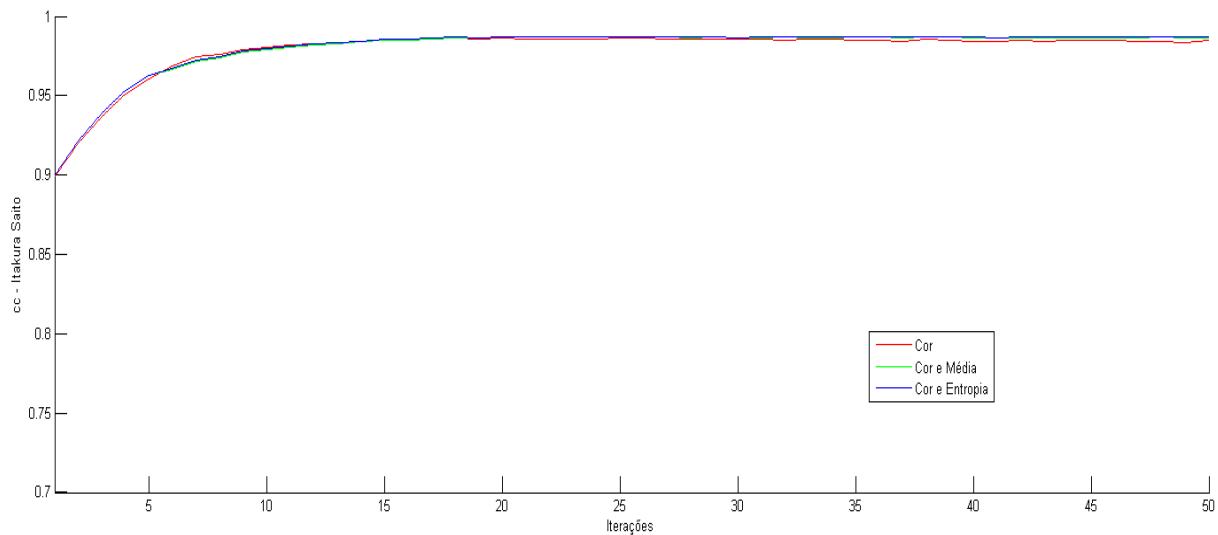


Figura 40 – Coeficiente de correlação- *Itakura Saito*

A distância *Squared loss*, *Generalized I-divergence* e *Logistic loss* destacaram-se nos modelos que utilizam “cor e média” e “cor”, mas foram menos eficazes quando utilizada a entropia, Figuras 41, 42 e 43. A agregação de características, neste experimento, não implicaram em maior eficácia dos modelos de registro propostos.

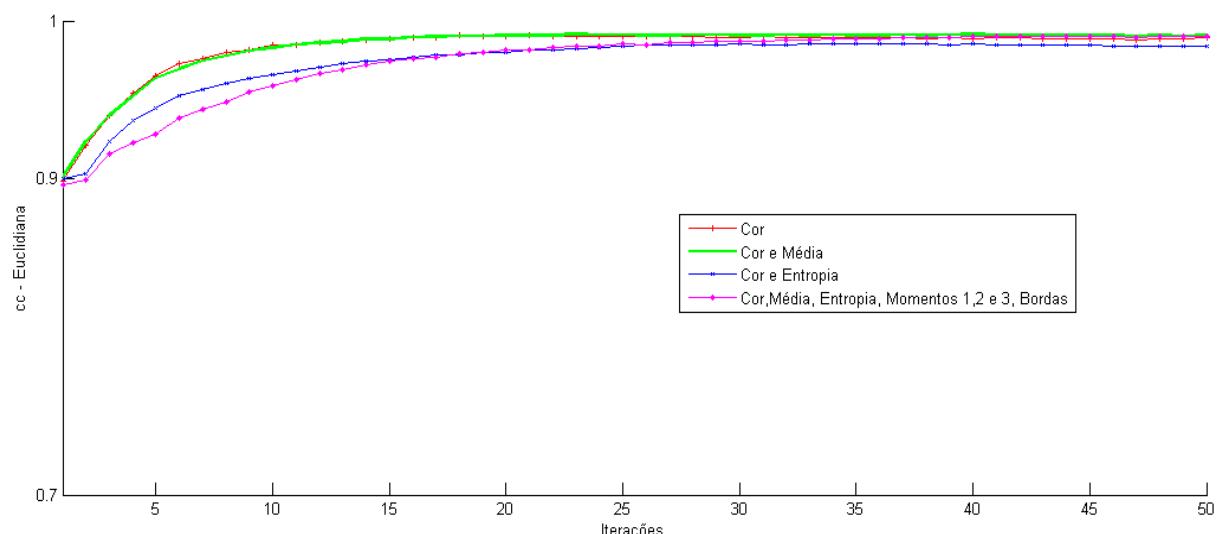
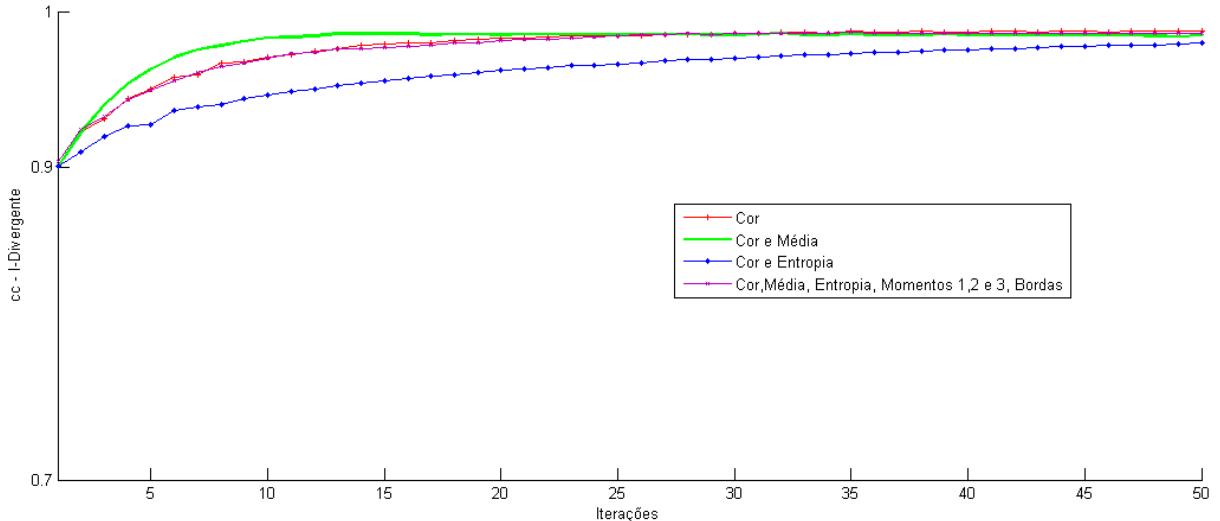
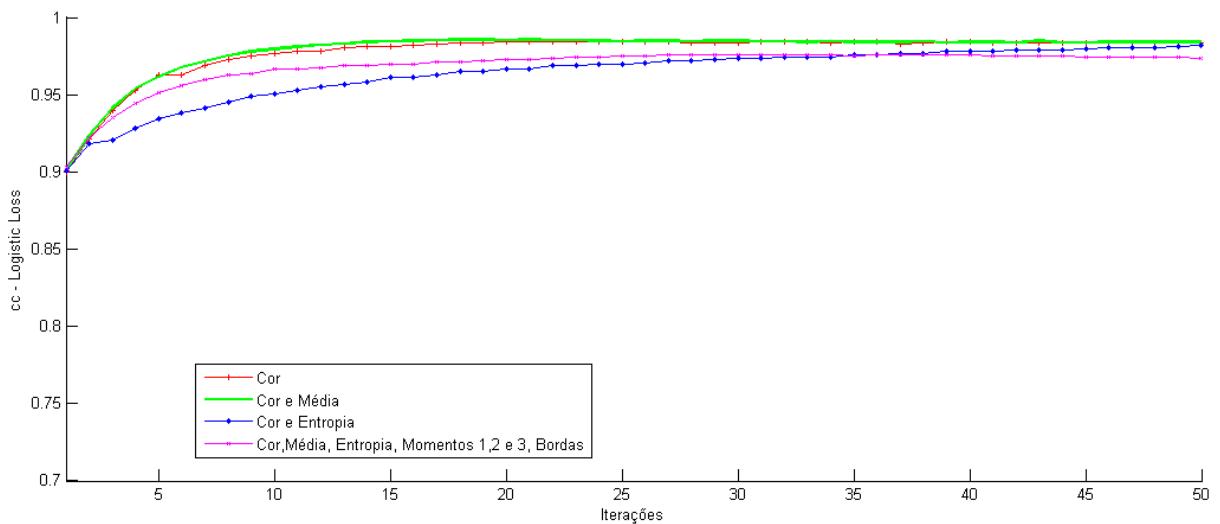


Figura 41 – Coeficiente de correlação- *Squared loss*

Figura 42 – Coeficiente de correlação- *Generalized I-divergence*Figura 43 – Coeficiente de correlação - *Logistic loss*

6.3.2.2 Experimento II

Nesta seção serão apresentados e analisados o registro realizado aplicando o modelo proposto às imagens, S e T, apresentadas na Figura 44(a) e(b), respectivamente. O coeficiente de correlação $CC(S, T)$ é igual 0,5856 e mede a diferença inicial entre estas imagens a qual pode ser visualizada na figura dada.

É importante observar que as figuras registradas têm formatos e cores diferentes. Assim, a imagem final $S(x + u(x))$ deve ter o formato da imagem modelo T e manter a cor dada inicialmente em S. Considere que S^{sl}, S^{ll}, S^{gid} e S^{is} representam as imagens $S(x + u(x))$ para os modelos de registro que utilizam, respectivamente, as divergências: *Squared loss*, *Logistic loss* e *Generalized I-divergence*.

Os testes foram realizados considerando diferentes combinações de características para as divergências propostas. A Tabela 10 resume os resultados obtidos considerando 200 iterações.

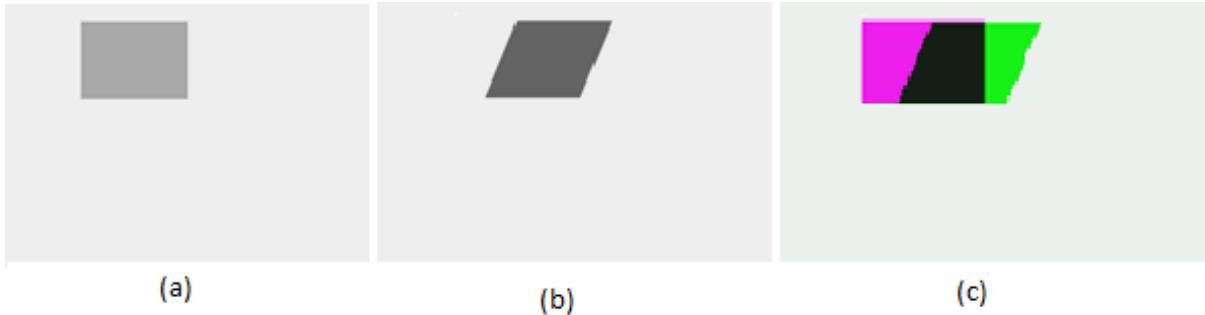


Figura 44 – (a) Imagem de referência S, (b) Imagem modelo T e (c) Diferença inicial entre S e T

Tabela 10 – Coeficientes de correlação obtidos para o registro das imagens S e T utilizando 200 iterações

Divergência	Caracterização	CC	Aumento relativo (%)
<i>Squared loss</i>	cor e entropia	0,9829	67%
<i>Squared loss</i>	cor e M2, M3, M4	0,9667	65%
<i>Squared loss</i>	cor e média	0,9653	65%
<i>Squared loss</i>	media e M2, M3, M4	0,9543	63%
<i>Gen. I-divergence</i>	cor e M2, M3, M4	0,9587	64%
<i>Gen. I-divergence</i>	cor e média	0,9522	63%
<i>Gen. I-divergence</i>	media e M2, M3, M4	0,9385	60%
<i>Gen. I-divergence</i>	borda e M2, M3, M4	0,9560	63%
<i>Logistic loss</i>	cor e entropia	0,9875	69%
<i>Logistic loss</i>	cor; M2, M3, M4	0,9357	60%
<i>Logistic loss</i>	cor e média	0,9644	65%
<i>Logistic loss</i>	média e M2, M3, M4	0,9357	60%
<i>Logistic loss</i>	borda e M2, M3, M4	0,8957	53%

M2,M3 e M4 representam, respectivamente, caracterizações obtidas calculando os momentos 2, 3 e 4 em cada pixel considerando oito vizinhos.

Tem-se que, para os testes realizados, o melhor desempenho para divergência *Squared loss* e *Generalized I-divergence* foi obtido com a caracterização “cor, M2,M3 e M4” com $CC(S^{sl}, T)=0,9667$ e $CC(S^{gid}, T)=0,9587$, respectivamente. O modelo de registro que aplica a divergência *Logistic loss* e “cor e entropia”, foi mais eficaz.

A Figura 45 demonstra estes resultados. Na primeira coluna tem a imagem modelo T representada em todas as linhas. Na segunda coluna temos as diferenças $S^{sl} - T$, $S^{gid} - T$ e $S^{ll} - T$, e na terceira tem-se as imagens finais S^{sl} , S^{gid} e S^{ll} .

Conforme observado no experimento I Figura 36 a imagem final, obtida empregando o modelo com a divergência *Logistic loss* associada à caracterização “cor e entropia”, era homogênea, mas incompleta. Entretanto neste experimento, o aumento do número de iterações possibilitou maior eficácia desta combinação quando comparada à todos os testes realizados.

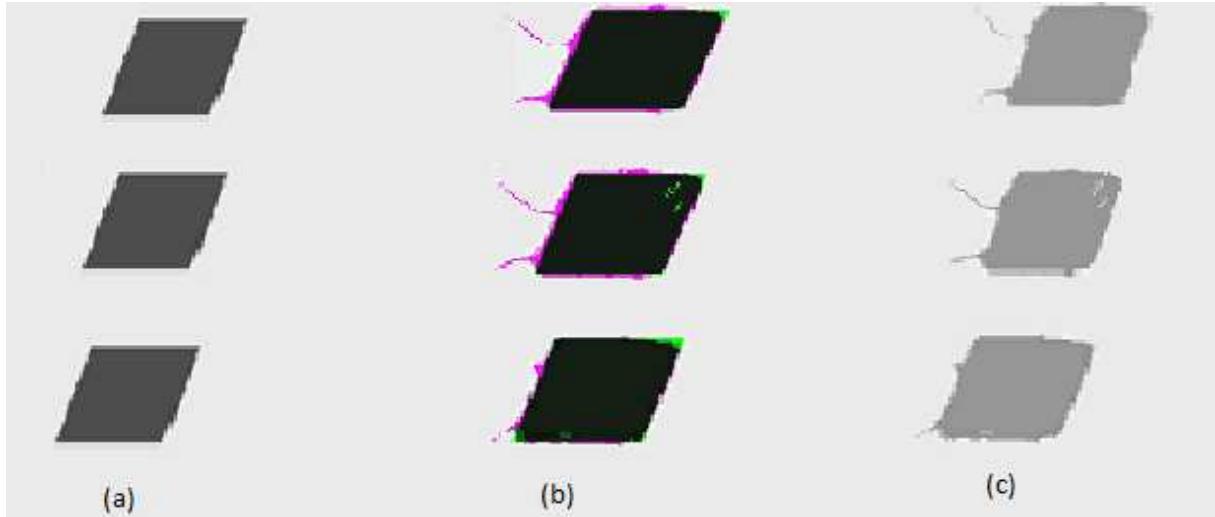


Figura 45 – Melhores resultados para as divergências *Squared loss*, *Generalized I-divergence* e *Logistic loss*. (a) imagem modelo T, (b) diferença entre a imagem T e as imagens finais, $S(x + u(x))$ e (c) imagens finais.

O número de iterações realizadas neste experimento, 200, é superior ao empregado anteriormente. O aumento do número de iterações também possibilitou um melhor desempenho desta divergência associada a caracterização “cor e média”, o que pode ser verificado na Tabela 11 a qual apresenta os coeficientes de correlação determinados utilizando 500 iterações.

Tabela 11 – Coeficientes de correlação obtidos para o registro das imagens S e T utilizando 500 iterações

Divergência	Caracterização	CC	Aumento relativo (%)
<i>Squared loss</i>	cor e média	0,9445	61%
<i>Generalized I-divergence</i>	cor e média	0,9623	64%
<i>Logistic loss</i>	cor e média	0,9652	65%

6.3.2.3 Considerações finais sobre a convergência do método proposto para diferentes opções de “Divergência” e “Caracterização”.

Nos experimentos I e II observamos as diferentes associações de “Divergência” e “Caracterização” podem proporcionar maior velocidade na convergência e tornar o método mais eficaz. A diferença obtida nos registros, ao modificarmos a escolha das divergências e ao utilizarmos outra caracterização além da “cor”, é mais evidente nas primeiras iterações do método.

Os gráficos apresentados ilustram a convergência dos métodos propostos com as diferentes associações. A convergência pode ser verificada observando a evolução do CC com relação às iterações. Entretanto, em geral, os modelos de registro obtidos associando as

divergências *Logistic loss*, *Generalized I-divergence* ou *Itakura Saito* com outras caracterizações além da “cor” propiciaram melhores resultados, quando comparados ao modelo tradicional definido com a divergência *Squared loss* e a caracterização “cor”.

Nas próximas seções apresentamos uma análise da interação entre os fatores “Divergência” e “Caracterização” e o efeito que estes exercem, individualmente, na qualidade do registro.

6.3.3 Análise da influência do fator “Divergência” no registro das imagens.

Nesta seção, no experimento III, avaliamos a influência isolada do fator “Divergência” na qualidade do registro das imagens e demonstramos que um melhor desempenho pode ser obtido alterando a divergência escolhida.

6.3.3.1 Experimento III

O modelo de registro proposto na equação (64) foi aplicado aos cinco pares de imagens médicas, apresentadas na Figura 46 a fim de avaliar a influência do fator “Divergência” no registro das imagens considerando para o fator “Caracterização” apenas a informação “cor”. O método proposto foi aplicado utilizando as divergências: *Squared loss*, *Logistic loss*, *Generalized I-divergence*.

As colunas da Figura 46 apresentam os resultados obtidos para cada par de imagens tomando as divergências *Squared loss*, *Logistic loss*, *Generalized I-divergence*. Os registros foram obtidos utilizando 500 iterações com passo $\Delta t = 0,07$ e $\lambda = 30$.

As diferenças entre as imagens, antes e após o registro, são representadas pelas cores verde e rosa. A terceira coluna apresenta a diferença inicial entre as imagens S e T. Apos a realização do registro, usando as quatro divergências propostas, ocorre uma redução desta diferença. Este fato pode ser verificado observando a atenuação das cores nas imagens das colunas 4, 5, 6 e 7.

A Tabela 12 apresenta a medida de avaliação CC após o registro, nas últimas colunas verifica-se a variação destes coeficientes em relação ao coeficiente de correlação inicial. Os coeficientes de correlação iniciais, $CC(S, T)$, para as imagens médicas: próstata, coração, peito, abdômen (corte 1) e abdômen (corte 2) são 0,3317, 0,8900, 0,9419, 0,7421 e 0,7514, respectivamente.

Os valores dos CC demonstram que o método proposto obteve um bom desempenho independente da escolha da divergência. Nos registros realizados houve aumento do CC em relação ao $CC_{inicial}$ e as imagens finais de 80% desses registros tem coeficientes de correlação maiores que 0,9109. Tem-se ainda que, em média, o método proposto propiciou um aumento relativo do coeficiente de correlação de aproximadamente 68% em relação aos coeficientes de correlação iniciais.

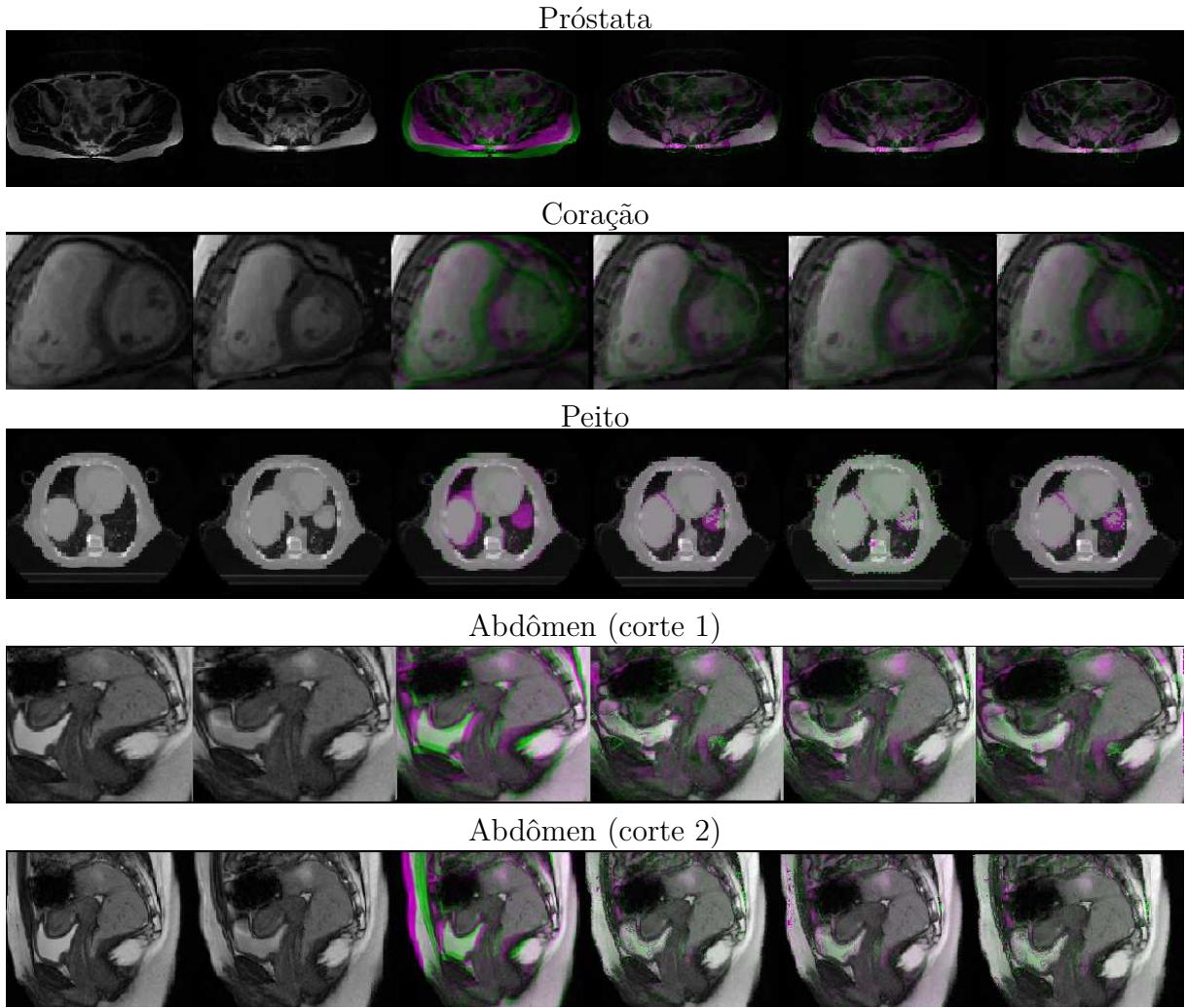


Figura 46 – Nas linhas tem-se as imagens S, T, a diferença inicial $S(x)-T(x)$, a diferença final $S^{eucl}(x+u)-T(x)$ obtida com a divergência *Squared loss*, a diferença final $S^{ll}(x+u)-T(x)$ obtida com a divergência *Logistic loss* e a diferença final $S^{id}(x+u)-T(x)$ considerando a divergência *Generalized I-divergence* para as imagens médicas: próstata, coração, peito, abdômen (corte 1) e abdômen (corte 2).

Tabela 12 – Medidas de avaliação obtidas nos registros das imagens.

	$CC_{inicial}$	CC_{final}			$CC_{rel} (%)$		
		Sq. loss	Log. loss	GID	Sq. loss	Log. loss	GID
Prostáta	0,3317	0,7868	0,8774	0,8724	68,1%	81,7%	80,9%
Coração	0,8900	0,9303	0,9533	0,9512	36,6%	57,5%	55,6%
Peito	0,9419	0,9753	0,9786	0,9827	57,1%	63,2%	70,2%
Abdômen (corte 1)	0,7421	0,9109	0,9187	0,9219	65,5%	68,5%	69,7%
Abdômen (corte 2)	0,7514	0,9496	0,9638	0,9532	79,7%	85,4%	81,2%
Média					61,4%	71,2%	71,5%

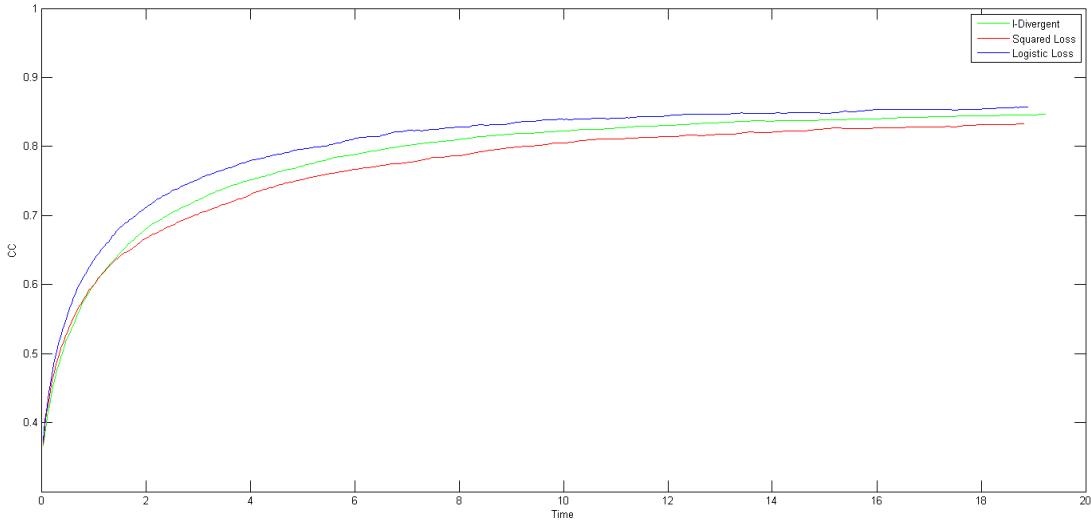
Tempo como critério de parada

A Tabela 13 apresenta uma comparação entre os tempos de processamento gasto nos registros do experimento. A diferença de tempo para as diferentes divergências é mínima.

Tabela 13 – Tempo de processamento do método proposto utilizando 500 iterações.

	Size	Tempo (min)		
		Sq. loss	Log. loss	GID
Próstata	256×256	18,8	18,9	19,2
Coração	100×100	3,0	3,5	3,8
Peito	100×100	4,0	3,0	3,1
Abdômen (corte 1)	170×170	10,1	10,0	9,6
Abdômen (corte 2)	190×240	12,8	13,4	13,1

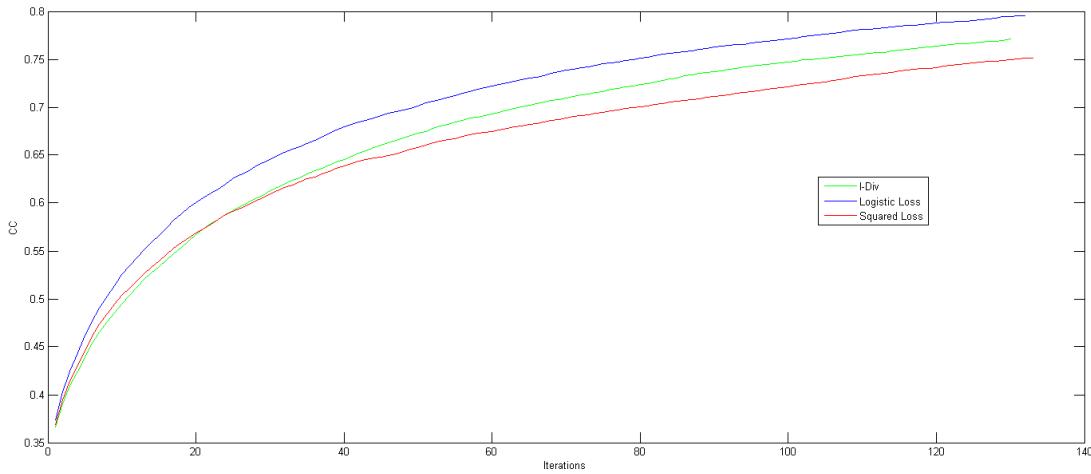
Para todas as escolhas de divergências, considerando os valores do parâmetro λ adotados, temos que o método proposto é convergente. A convergência pode ser verificada observando a evolução do CC no tempo. A Figura 47 apresenta a evolução do CC resultante do registro do par de imagens próstata considerando as divergências *Squared loss*, *Logistic loss* e *Generalized I-divergence*. A Figura 48 apresenta a evolução do CC em relação ao número de iterações após 5 minutos de processamento. As divergências *Logistic loss* e *Generalized I-divergence* apresentam, respectivamente, coeficientes de correlação iguais a 0,7960 e 0,7709 enquanto que o registro com a *Squared loss* alcançou CC igual a 0,7516.

Figura 47 – Evolução do CC resultante do registro do par de imagens próstata considerando as divergências *Squared loss*, *Logistic loss* e *Generalized I-divergence*.

6.3.3.2 Considerações finais sobre a influência do fator “Divergência”.

Na Figura 49(a) pode-se visualizar a evolução do CC para cada par de imagem considerando as diferentes divergências. Os CC obtidos para as divergências *Logistic loss* e *Generalized I-divergence* são maiores que os obtidos com a *Squared loss* em todos os registros.

Figura 48 – Evolução do CC em relação ao número de iterações após 5 minutos de processamento.



O efeito da escolha da divergência de Bregman considerando os registros realizados pode ser visualizado no gráfico de efeito apresentado na Figura 49(b) obtido a partir das médias dos CC_{rel} de cada nível do fator “Divergência”. Observe que a escolha de diferentes divergências de Bregman influencia no aumento relativo do coeficiente de correlação.

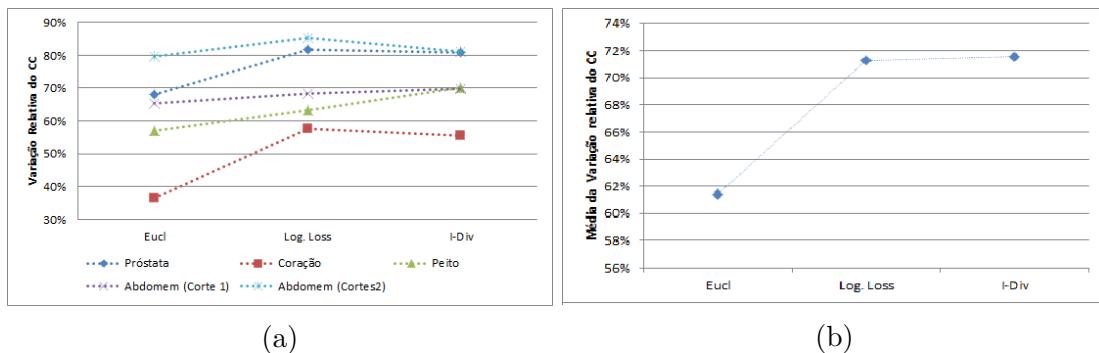


Figura 49 – (a) Evolução da variação do CC para cada par de imagem considerando as diferentes divergências. (b) Efeito do fator “Divergência” sobre o aumento relativo coeficiente de correlação das imagens registradas considerando a caracterização “cor”

A escolha das divergências *Logistic loss* e *Generalized I-divergence* proporcionaram em média, aproximadamente, um aumento relativo do CC, 10% maior que o aumento proporcionado pela escolha da divergência *Squared loss*.

Estes resultados, considerando o número de iterações utilizado no processo, indicam que a escolha das divergências *Logistic loss* e *Generalized I-divergence* como medidas de similaridade tornaram o método de registro mais eficaz que o método tradicional cuja medida de similaridade é dada pela divergência *Squared loss*. Desta forma, constata-se que o fator “Divergência” tem impacto na eficácia do registro da amostra de imagens.

6.3.4 Análise do impacto do parâmetro lambda.

No experimento III constatou-se a influência do fator “Divergência” no processo de registro considerando o mesmo valor de balanceamento ($\lambda = 30$) para as diferentes escolhas de divergências avaliadas. Entretanto, a diferença entre os limites superiores dos conjuntos imagem das divergências apresentadas na Tabela 9 sugere que o valor do parâmetro λ deve ser diferente conforme a divergência utilizada no registro.

Para definirmos valores ideais desse parâmetro para as três divergências utilizadas nos registros das imagens apresentadas na Figura 46, propomos uma abordagem baseada na relação entre elas. Esta abordagem é apresentada no experimento IV. A relação é definida utilizando o método de regressão linear. A partir do valor ótimo para o registro com a divergência *Logistic loss*, definido experimentalmente, obtemos, através das relações estabelecidas, os outros dois parâmetros.

Pretende-se, desta forma, verificar o impacto do valor do termo de balanceamento λ nos resultados obtidos.

6.3.4.1 Experimento IV

Os pares de imagens S e T apresentados na primeira e segunda colunas da Figura 46 serão registrados utilizando para λ os valores: 5,15, 30, 50 e 60. Analogamente ao experimento anterior considerou-se, para esta análise, o fator “Caracterização” dado apenas pela informação “cor”, 500 iterações e passo $\Delta t = 0,07$. As Tabela 14 e 15 apresentam, respectivamente, os coeficientes de correlação final e as variações desta medida obtidos após o registro utilizando cada um dos valores de λ e divergência *Logistic loss*.

Tabela 14 – Coeficientes de correlação final para “Caracterização” cor e “Divergência” *Logistic loss*, 500 iterações, $\Delta t = 0,07$ e $\lambda = 5, 15, 30, 50$ e 60 .

	CC incial	CC_{final}				
		$\lambda = 5$	$\lambda = 15$	$\lambda = 30$	$\lambda = 50$	$\lambda = 60$
Próstata	0,3317	0,7562	0,8449	0,8774	0,8786	0,8781
Coração	0,8900	0,9275	0,9427	0,9533	0,813	0,786
Peito	0,9419	0,9709	0,9778	0,9786	0,9763	0,9616
Abdômen (corte 1)	0,7421	0,8864	0,9171	0,9185	0,9199	0,9195
Abdômen (corte 2)	0,7514	0,9275	0,954	0,9638	0,9608	0,9561

A Figura 50 ilustra a evolução do coeficiente de correlação para o par de imagens Abdômen (corte 2).

Os resultados destas tabelas indicam que a escolha do λ influencia na qualidade do registro podendo comprometer o processo. Este fato pode ser observado no registro do par de imagens Coração. As imagens finais obtidas no registro destas imagens, para os valores de $\lambda = 50$ e $\lambda = 60$, não apresentam crescimento do coeficiente de correlação. Este comportamento é representado pelo símbolo “–”.

Tabela 15 – Diferença entre os coeficientes de correlação inicial e final obtido considerando para o fator “Caracterização” a cor e para a “Divergência” a opção *Logistic loss*. 500 iterações, $\Delta t = 0,07$ e $\lambda = 5, 15, 30, 50$ e 60 .

	$CC_{final} - CC_{inicial}$				
	5	15	30	50	60
Próstata	0,425	0,513	0,546	0,547	0,546
Coração	0,038	0,053	0,063	-	-
Peito	0,029	0,036	0,037	0,034	0,020
Abdômen (corte 1)	0,144	0,175	0,176	0,178	0,177
Abdômen (corte 2)	0,176	0,203	0,215	0,209	0,205

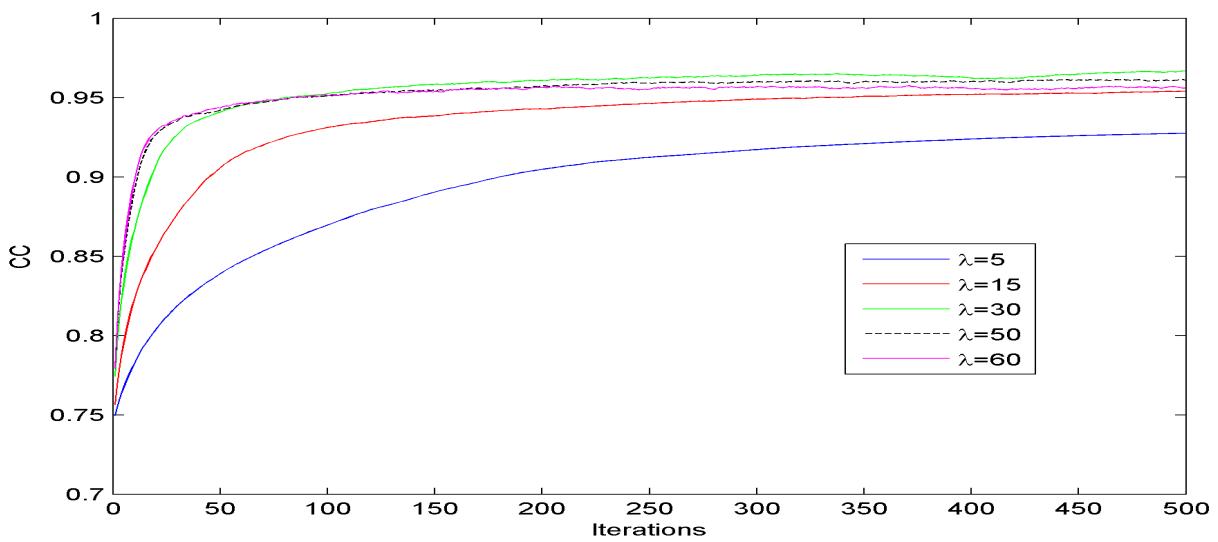


Figura 50 – Evolução do coeficiente de correlação para o par de imagens abdômen (corte 2).

O registro dos pares de imagem coração, peito e abdômen (corte 2), para os valores de $\lambda = 50$ e $\lambda = 60$, apresentam variação relativa do CC menores que os obtido para $\lambda = 30$. Sendo assim, dos valores considerados para parâmetro λ neste experimento, “ $\lambda = 30$ ” proporcionou, em média, o maior aumento relativo do coeficiente de correlação.

Observe que adotando $\lambda = 30$ tem-se em média uma maior eficácia no registro das imagens utilizadas neste trabalho quando a divergência *Logistic loss* é utilizada. A variação relativa média é 71,46% (Tabela 15).

A partir deste valor ótimo, definido experimentalmente, para a divergência *Logistic loss* podemos definir os valores do parâmetro para as outras duas divergências considerando a relação entre elas. As relações são obtidas utilizando o método estatístico de regressão linear.

As divergências *Logistic loss*, *Squared loss* e *Generalized I-divergence* foram aplicadas a 40 pares de imagens médicas incluindo entre elas os pares de imagens registradas no experimento III.

Aplicando o método de regressão linear temos que a divergência *Logistic loss* está

associada, respectivamente, com as divergências *Squared loss* e *Generalized I-divergence* pelas seguintes equações de regressão:

$$d_{ll}(S, T) = 3,1 d_{sl}(S, T), \quad (74)$$

$$d_{ll}(S, T) = 1,4 d_{Id}(S, T), \quad (75)$$

onde $d_{ll}(S, T)$, $d_{sl}(S, T)$ e $d_{Id}(S, T)$ são, respectivamente as divergências *Logistic loss*, *Squared loss* e *Generalized I-divergence*.

Os coeficientes de correlação de Pearson, $R = 0,95$ e $R = 0,98$, indicam uma alta correlação entre estes pares de divergências. A Figura 51 apresenta as relações e o quadrado dos coeficientes de correlação obtidos.

A partir destas relações podemos estabelecer os valores λ para as divergências *Squared loss* e *Generalized I-divergence*. Se λ_{ll} é o valor do parâmetro utilizado no registro com a *Logistic loss*, temos que:

$$\begin{aligned} \lambda_{ll} d_{ll}(S, T) &= 3,1 \lambda_{sl} d_{sl}(S, T), \\ \lambda_{sl} &= 3,1 \lambda_{ll}, \end{aligned} \quad (76)$$

onde λ_{sl} é o valor do parâmetro utilizado no registro com a *Squared loss*. Da equação (75), sendo λ_{id} o parâmetro para a *Generalized I-divergence*, segue que:

$$\lambda_{id} = 1,4 \lambda_{ll}. \quad (77)$$

Da abordagem apresentada temos que, quando a divergência *Logistic loss* é utilizada com $\lambda = 30$ tem-se, em média, uma maior eficácia no registro. Desta forma, com o intuito de analisarmos o efeito das divergências tomamos, $\lambda = 30, 90$ e 42 respectivamente, para os parâmetros utilizados com as divergências *Logistic loss*, *Squared loss* e *Generalized I-divergence*. A Tabela 16 apresenta o coeficiente de correlação e a variação relativa desta medida para cada registro realizado.

Tabela 16 – Medidas de avaliação obtidas no registro das imagens da Figura 46 com N.Iter.= 500, $\Delta = 0,07$, $\lambda = 90, 30$ e 42 .

	$CC_{inicial}$		CC_{final}			CC_{rel} (%)		
	Sq. loss	Log. loss	GID	Sq. loss	Log. loss	GID		
Próstata	0,3317	0,8524	0,8774	0,8743	77,9%	81,7%	81,2%	
Coração	0,8900	0,9495	0,9533	0,9524	54,1%	57,5%	55,7%	
Peito	0,9419	0,9751	0,9786	0,9828	57,1%	63,2%	70,4%	
Abdômen (corte 1)	0,7421	0,9178	0,9187	0,9230	68,1%	68,5%	70,1%	
Abdômen (corte 2)	0,7514	0,9556	0,9638	0,9565	82,1%	85,4%	82,5%	
Média					67,88%	71,26%	72,19%	

Os resultados demonstram que os métodos com as divergências *Squared loss* e *Generalized I-divergence* obtiveram melhores resultados considerando os novos parâmetros

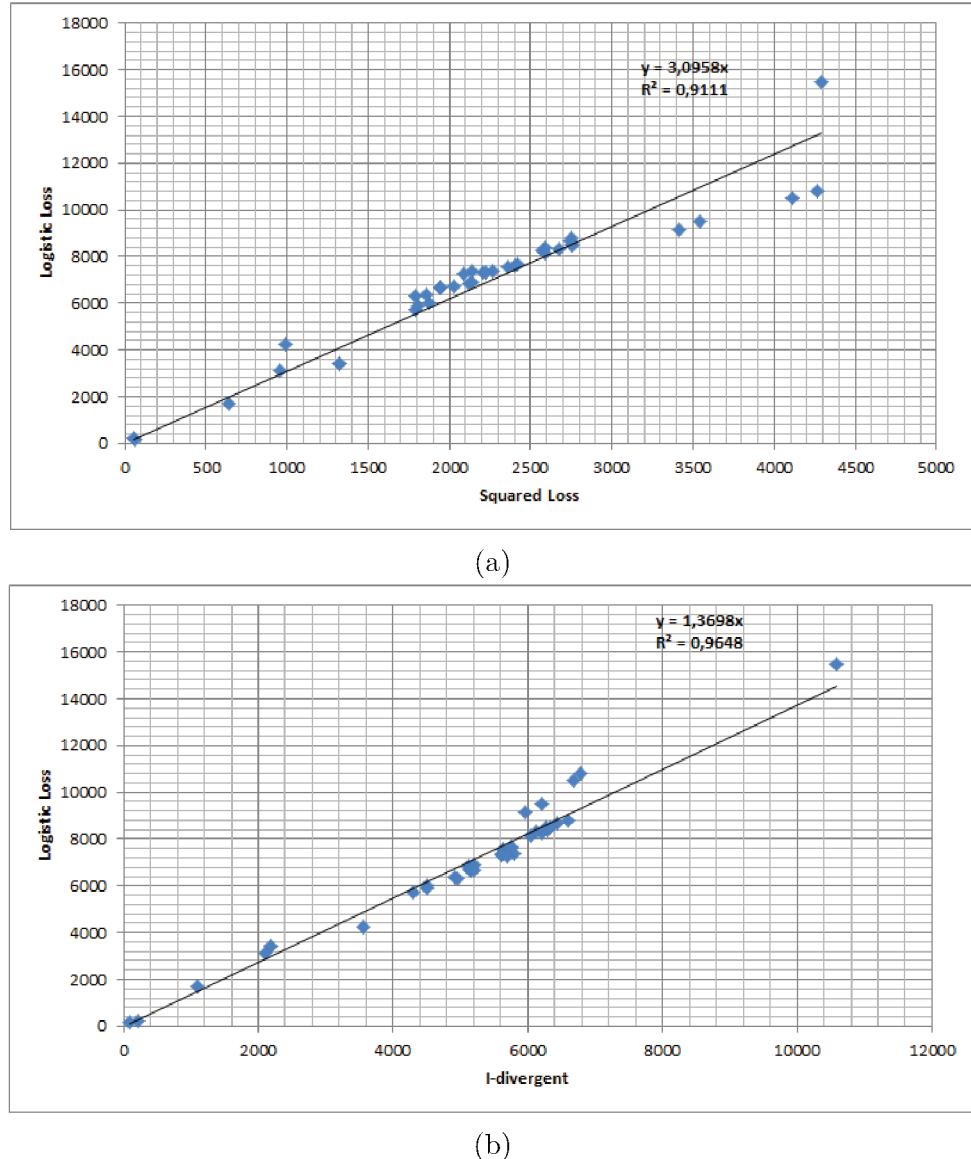


Figura 51 – (a) Divergência *Logistic loss* em função da divergência *Squared loss* (b) Divergência *Logistic loss* em função da divergência *Generalized I-divergence*.

adotados. Destacamos o aumento do CC obtido nos registros utilizando a *Squared loss*, o que proporcionou uma variação de aproximadamente 5% na média do CC_{rel} , veja Tabela 12. Embora os resultados considerando os novos parâmetros sejam melhores temos ainda uma maior eficácia do método com as divergências *Logistic loss* e *Generalized I-divergence*.

6.3.4.2 Considerações finais sobre o impacto do parâmetro λ .

Maiores valores para o parâmetro de balanceamento podem propiciar maior velocidade na convergência. Entretanto, em nosso modelo λ multiplica o termo de similaridade. Desta forma, valores altos para este parâmetro devem ser assumidos com cuidado pois podem tornar insignificante o efeito do termo de regularização. Comprometendo, assim, a suavização do campo de deformação obtido durante o processo.

6.3.5 Análise do efeito causado pela agregação de características

O modelo variacional proposto permite que sejam incorporadas simultaneamente diferentes informações locais ao processo de registro. Nesta seção tem-se como objetivo avaliar o efeito provocado pela associação de informações locais na qualidade do registro. Apresentam-se registros considerando diferentes opções para os fatores “Caracterização” e “Divergência” evidenciando a relevância da utilização de informações locais no procedimento de registro. Pretendeu-se, assim, verificar se, de fato, a agregação de informações possibilita um melhor desempenho do processo de registro.

Inicialmente, no experimento V, buscamos avaliar o impacto que uma informação isolada causaria no procedimento de registro. A característica escolhida para esta análise foi a “entropia”. Em seguida foi realizado uma comparação do registro de imagens considerando, para o fator “Caracterização”, a associação “cor e entropia” com o registro obtido utilizando as informações isoladas “cor” e “entropia.”.

Em seguida no experimento VI analisou-se o efeito causado pela agregação de informações no processo de registro, comparando as associações “cor”, “cor e entropia”, “cor, entropia e desvio” e “cor, entropia, desvio e média”. A ideia aqui foi avaliar a qualidade do registro na medida em que novas informações fossem incorporadas ao processo. A interação entre os fatores “Caracterização” e “Divergência”, bem como o efeito de cada um deles na qualidade do registro, foram analisadas.

Finalmente, no experimento VI foi avaliado o registro, utilizando como fator de “Caracterização”, combinações obtidas com os momentos invariantes 1 e 2 de “Hu”, (FLUSSER, 2006).

6.3.5.1 Experimento V

Neste experimento aplicou-se o modelo de registro proposto na equação (64) para o par de imagens médicas S e T cuja diferença inicial é apresentada na 4^a linha da Figura 46. O método proposto foi avaliado considerando o fator “Divergência” representado pela *Squared loss*, *Logistic loss* e *Generalized I-divergence*. A fim de avaliar a influência da informação “entropia” no registro das imagens dadas utilizou-se inicialmente, para o fator “Caracterização”, apenas a “entropia”, em seguida, considerou-se as opções “cor” e “cor e entropia”

A Figura 52 apresenta as imagens S e T representadas pela “entropia” e a diferença entre elas.

Para cada combinação da característica “entropia” e divergências foram realizadas 1000 iterações com passo $\Delta t = 0,07$ e $\lambda = 50$. O coeficiente de correlação inicial entre as imagens S e T, $CC(S, T)$, considerando a caracterização dada pela entropia, é igual a 0,7827.

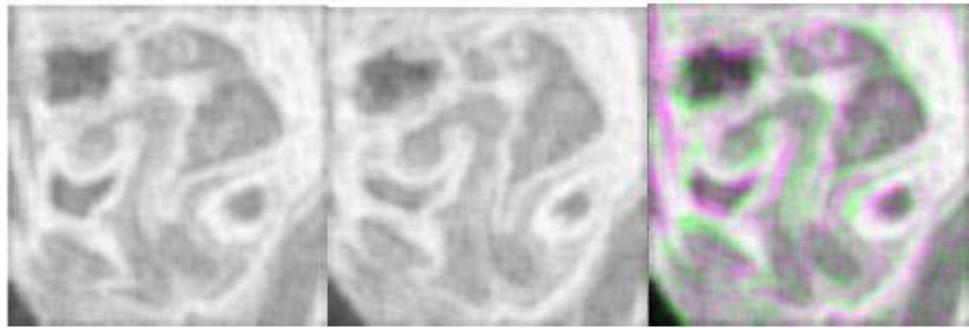


Figura 52 – Imagens S, T e diferença inicial caracterizadas pela “entropia”.

Tabela 17 – Medidas de avaliação obtidas considerando as imagens representadas pela “entropia”- $N^oIter = 1000$, $\Delta t = 0,07$ e $\lambda = 50$.

Divergência	CC_{final}	RMSE	$CC_{rel}(\%)$
<i>Squared loss</i>	0,9455	0,0122	74,9%
<i>Logistic loss</i>	0,9649	0,0118	83,8%
<i>Generalized I-divergence</i>	0,9419	0,0150	73,3%

A Tabela 17 apresenta a avaliação do registro obtido utilizando apenas a informação entropia na caracterização das imagens S e T, a cor não foi considerada. A divergência *Logistic loss* apresentou o melhor desempenho propiciando um aumento relativo do $CC_{inicial}$ de 83,8%. Este valor é aproximadamente 10% maior que os CC finais dados pelas divergências *Squared loss* e *Generalized I-divergence*.

O registro das imagens representadas pela entropia foi muito eficaz. Entretanto na prática o registro final deve ser apresentado utilizando níveis de cinza. Assim, aplicamos o campo de deslocamento obtido nesse processo na imagem S caracterizada pela matriz “cor”.

Os resultados são apresentados na Tabela 18. O coeficiente de correlação inicial entre as imagens S e T, $CC(S, T)$, é igual a 0,7421.

Tabela 18 – Medidas de avaliação aplicando o campo de deformação u obtido pela caracterização “entropia” às imagens iniciais S e T - $N^oIter = 1000$, $\Delta t = 0,07$ e $\lambda = 50$.

Divergência	CC_{final}	RMSE	$CC_{rel}(\%)$
<i>Squared loss</i>	0,8502	0,0437	41,9%
<i>Logistic loss</i>	0,8534	0,0431	43,1%
<i>Generalized I-divergence</i>	0,8263	0,0472	32,6%

Os campos de deslocamento obtidos, considerando as três divergências, propiciaram um aumento relativo do $CC_{inicial}$. O menor aumento foi resultado do registro com a *Generalized I-divergence*, de 32,6%. Estes resultados são piores quando comparados aos

obtidos considerando o campo de deslocamento resultante do registro cuja caracterização é dada somente pela cor, conforme apresentado na tabela 12.

Embora isoladamente a caracterização entropia propicie um menor aumento dos CC das imagens, a associação desta característica com a cor torna o método mais eficaz. Estes resultados podem ser observados na Tabela 19 a qual apresenta uma comparação dos valores obtidos considerando “cor” e “cor e entropia” considerando 500 iterações, $\Delta t = 0,07$ e $\lambda = 30$.

Tabela 19 – Comparação dos coeficientes de correlação obtidos considerando “cor” e “cor e entropia - $N^o Iter = 500$, $\Delta t = 0,07$ e $\lambda = 30$.¹

	“CC- cor”	CC - “cor e entropia”	Contribuição da “entropia”
<i>Squared loss</i>	0,9109	0,9227	0,012
<i>Logistic loss</i>	0,9187	0,9344	0,016

Na Tabela 19 temos uma comparação entre os CC finais obtidos com a cor e com a associação desta característica com a entropia. Esta comparação é realizada a fim de identificar o efeito desta associação no resultado final. A divergência *Logistic loss* foi escolhida para esta comparação pois apresentou maior eficácia quando consideramos isoladamente a caracterização entropia. A última coluna desta tabela apresenta a diferença entre os resultados considerando a caracterização cor e a associação cor e entropia. Em ambos os casos a entropia propiciou um aumento do CC final.

Constata-se que a associação de informações, de fato, proporcionou um melhor desempenho no registro, quando comparada aos resultados obtidos considerando-as isoladas. A divergência *Logistic loss* proporciona uma maior contribuição com associação “cor e entropia” do que a distância *Squared loss*, embora esta, também apresente um alto desempenho com as duas caracterizações. Os resultados apresentados indicam a habilidade da divergência *Logistic loss* no registro de imagens quando o fator “Caracterização” envolve a “entropia”.

A Figura 53, destaca a diferença inicial entre S e T, as diferenças finais entre $S(x+u(x))$ e T e as imagens finais, $S(x + u(x))$, considerando, respectivamente, as caracterizações, “cor”, “entropia” e “cor e entropia”, para as divergências *Squared loss* e *Logistic loss*.

6.3.5.2 Experimento VI

Neste experimento aplicou-se o modelo de registro proposto na equação (64) para o par de imagens S e T cuja diferença inicial é apresentada na primeira linha da Figura 46. Para cada combinação de característica e divergência foram realizadas 500 iterações com passo $\Delta t = 0,07$ e $\lambda = 30$. O coeficiente de correlação incial entre as imagens S e T, $CC(S, T)$, é igual a 0,3317.

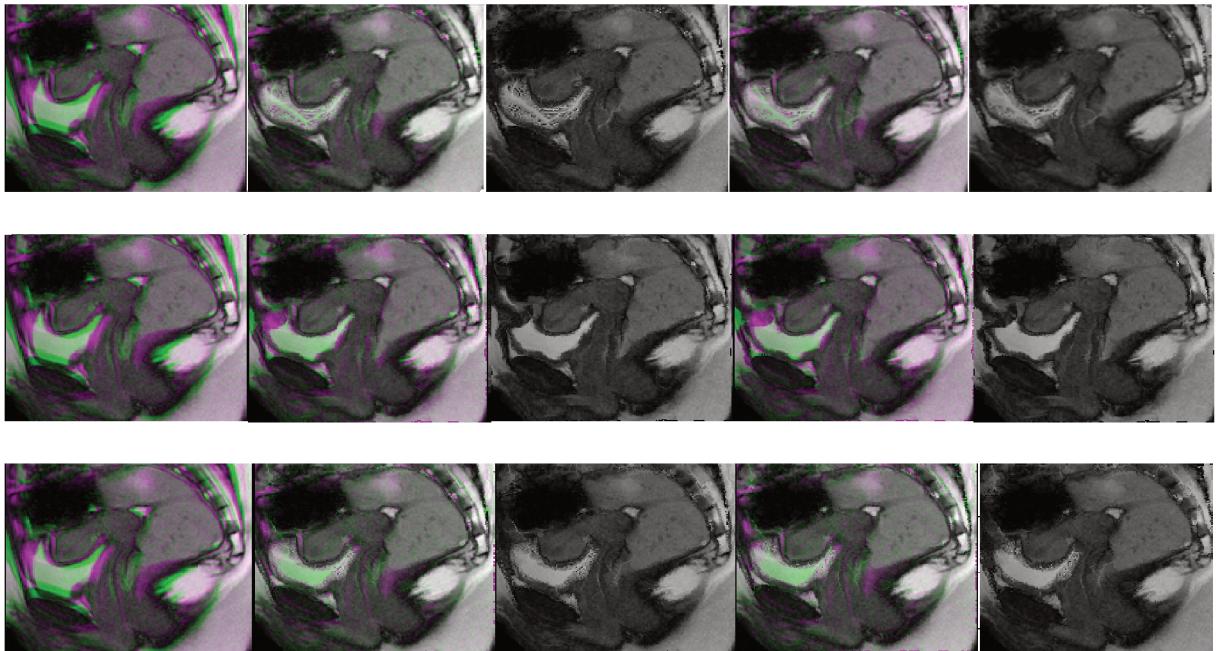


Figura 53 – Diferença inicial entre S e T , diferença final entre $S(x + u(x))$ e T e imagem final considerando a divergência *Squared loss*, diferença final entre $S(x+u(x))$ e T e imagem final considerando a divergência *Logistic loss*. Resultados para as caracterizações “cor”(1^a linha), “entropia”(2^a linha) and ‘cor and entropia’ (3^a) linha.

A Tabela 20 apresenta a variação relativa do coeficiente de correlação obtido para cada um dos registros realizados. A combinação “cor, entropia, desvio e média” apresenta os melhores resultados em todas as opções do fator “Divergência”. As imagens finais obtidas considerando as três opções deste fator são apresentadas na Figura 54.

Tabela 20 – Variação relativa do coeficiente de correlação - $N^o Iter = 500$, $\Delta t = 0,07$, $\lambda = 30$ e $CC_{inicial} = 0,3317$.

	CC_{final}			$CC_{rel} (%)$		
	Sq. loss	Log. loss	GID	Sq. loss	Log. loss	GID
Cor	0,7868	0,8774	0,8724	68,1%	81,7%	80,9%
Cor e entropia	0,8335	0,8849	0,8886	75,1%	82,8%	83,3%
Cor, entropia e desvio	0,8741	0,8940	0,8931	81,2%	84,1%	84,0%
Cor, entropia, desvio e média	0,9000	0,9120	0,9234	85,0%	86,8%	88,5%

Todas as associações propostas quando comparadas com a configuração tradicional, “cor” e divergência “*Squared loss*”, propicia um melhor desempenho do registro, sendo o melhor resultado obtido combinando a divergência *Generalized I-divergence* com a caracterização “cor, entropia,desvio e média”. A divergência *Generalized I-divergence* utilizada com a caracterização “cor, entropia, desvio padrão e média” propiciou um CC_{final} de 0.923 que é 20,4% maior que ao CC_{final} obtido utilizando a configuração tradicional: divergência *Squared loss* e cor.

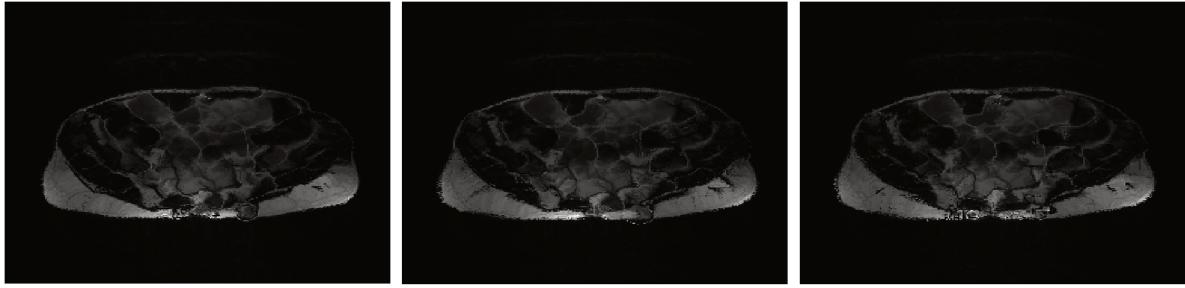


Figura 54 – Resultado caracterização “cor, entropia, desvio e média”: imagem final considerando as divergências: *Squared loss*, *Generalized I-divergence* e *Logistic loss*, respectivamente.

A Tabela 21 apresenta a contribuição de cada configuração em relação a esta configuração tradicional, dada por:

$$\text{Contribuição}_{\text{cor}+\text{Eucl}} = CC_{\text{rel}} - 68,1\%.$$

A divergência *Squared loss* foi mais afetada pela agregação de características que outras

Tabela 21 – Contribuição obtida em cada combinação em relação ao método tradicional: “*Squared loss* e cor” - *Iterations* = 500, Δt = 0,07, λ = 30.

	Contribuição em relação à “ <i>Squared loss + cor</i> ”		
	Sq. loss	Log. loss	GID
Cor	0,0%	13,6%	12,8%
Cor e entropia	7,0%	14,7%	15,2%
Cor, entropia e desvio	13,1%	16,0%	15,9%
Cor, entropia, desvio e média	16,9%	18,7%	20,4%

divergências, embora o CC_{final} seja inferior em todos os casos. Ao agregar as informações “entropia, desvio padrão e média” à caracterização “cor” obteve-se um aumento relativo de 16,9%, considerando a divergência *Squared loss*. Nos casos das divergências *Logistic loss* e *Generalized I-divergence* o aumento relativo foi de 5,2% e de 7,6%, respectivamente. Este comportamento pode ser visualizado no gráfico de interação apresentado na Figura 55(a).

No gráfico de interação entre os fatores os segmentos não são paralelos o que nos permite concluir que existe mudança de comportamento de um fator para as diferentes escolhas do outro fator, com relação a variação do coeficiente de correlação.

A Tabela 22 apresenta as contribuições obtidas pela agregação das informações para cada uma das opções de divergência. Fixando a divergência tem-se o coeficiente de variação relativa para a configuração “cor” ($CC_{\text{rel}}^{\text{cor}}$) em cada coluna. Assim as contribuições, fixando as divergências, são dadas por:

$$\text{Contribuição}_{\text{cor}} = CC_{\text{rel}} - CC_{\text{rel}}^{\text{cor}}.$$

Tabela 22 – Contribuição obtida pela agregação de informação em relação à característica “cor” para as divergências: *Squared loss*, *Logistic loss*, *Generalized I-divergence*.

	Contribuição em relação a “cor”		
	Sq. loss	Log. loss	GID
Cor	-	-	-
Cor e entropia	7,0%	1,1%	2,4%
Cor, entropia e desvio	13,1%	2,5%	3,1%
Cor, entropia, desvio e média	16,9%	5,2%	7,6%

Os resultados comprovam que CC é afetado pela divergência utilizada. Esta influência é ilustrada na Figura 55(b). Quando a divergência é modificada o valor do coeficiente de correlação relativo é alterado.

A Figura 55(c) apresenta o efeito que a associação de características provocou no resultado final do registro. Este gráfico é obtido considerando média das variações relativas para cada caracterização. A associação de características provoca, em média, um crescimento de aproximadamente 10% no $CC_{relative}$.

Os gráficos de efeito principal apresentados nas Figuras 55(b) e 55(c) evidenciam que tanto as informações: cor, entropia, desvio e média, quanto a escolha da divergência

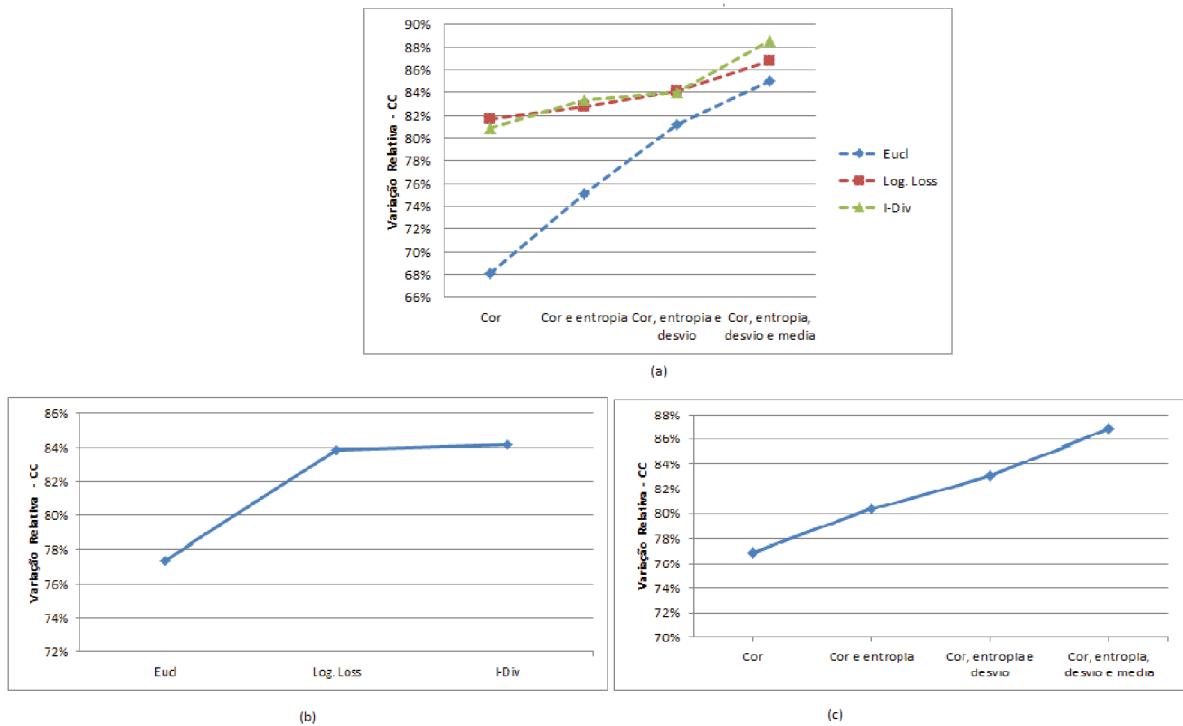


Figura 55 – (a) Evolução da variação do CC para cada par de imagem considerando as diferentes caracterizações. (b) Efeito do fator “Caracterização” sobre a variação do coeficiente de correlação. (c) Efeito do fator “Divergência” sobre a variação relativa do coeficiente de correlação.

influenciam a qualidade do registro.

6.3.5.3 Experimento VII

Em seguida, o impacto da agregação de informações, é avaliado utilizando a informação “momentos”, composta pelos momentos invariantes 1 e 2 de “Hu”.

Os registros apresentados neste experimento visam demonstrar a influência que os momentos estatísticos tem sobre o processo. Os pares de imagens apresentadas nas duas primeiras colunas da 1^a, 2^a e 5^a linhas da Figura 46 foram registradas definindo para o fator “Divergência” os níveis *Logistic loss* e *Squared loss*. Para o fator “Caracterização” além da “cor” utilizou-se a informação “momentos” composta pelos momentos invariantes 1 e 2 de “Hu” calculados em cada pixel considerando uma vizinhança 8x8. O funcional proposto é aplicado considerando duas informações, ou seja, duas matrizes. As imagens apresentadas na Figura 56 são exemplos da aplicação dessa caracterização na imagem de uma próstata.

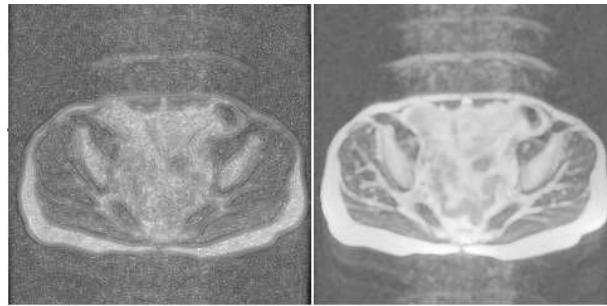


Figura 56 – Caracterização “momentos”: invariante 1 e invariante 2.

Para cada combinação de característica e divergência foram realizadas 500 iterações com passo $\Delta t = 0,07$ e $\lambda = 30$. A Tabela 23 apresenta a variação relativa do coeficiente de correlação obtido para cada um dos registros realizados.

Tabela 23 – Variação relativa do coeficiente de correlação obtido para os registros realizados, considerando as informações cor e momentos.

	<i>Squared loss</i>			<i>Logistic loss</i>		
	Próstata	Coração	Abdômen	Próstata	Coração	Abdômen
Cor e momentos	75,7%	43,2%	82,2%	84,3%	58,3%	85,0%
Cor	68,1%	36,6%	79,7%	81,8%	57,5%	85,4%

Da análise dos resultados percebe-se que a atribuição de informações locais através dos momentos invariantes tem um efeito positivo no registro das imagens, considerando tanto a distância *Squared loss* quanto a divergência *Logistic loss*. A Figura 57(b) apresenta o efeito principal desta informação nos registros realizados. Observe que ao agregar a característica “momentos” ao processo de registro tem-se em média um aumento de

aproximadamente 2% na variação do coeficiente de correlação das imagens. A Figura 57(a) representa a interação entre os fatores “Caracterização” e “Divergência”. O melhor resultado entre os registros realizados foi obtido utilizando a divergência *Logistic loss* e as informações de cor e momentos. As retas inclinadas observadas no gráfico evidenciam a interação entre os fatores e demonstram que o efeito provocado pelo fator “Divergência” depende da opção “Caracterização”.

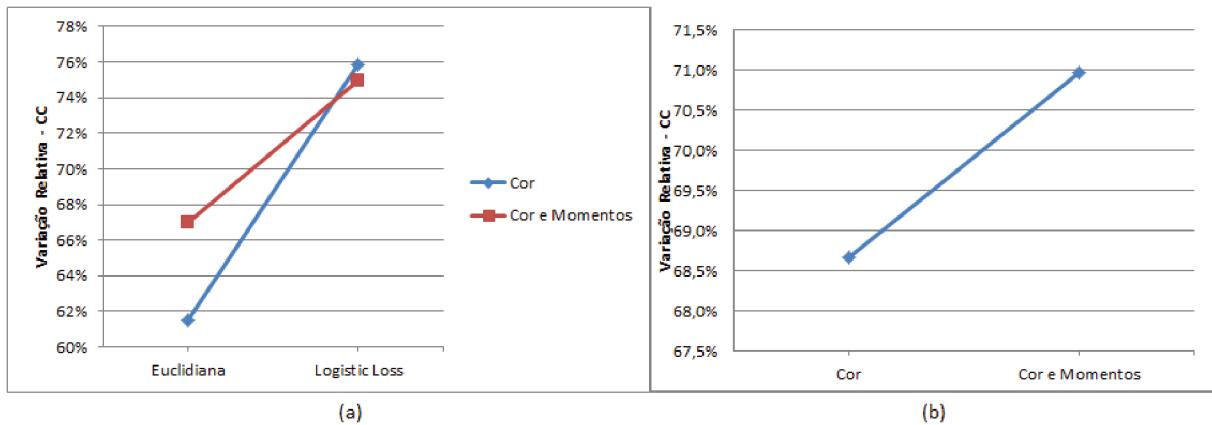


Figura 57 – (a) Interação entre a “Caracterização” e “Divergência”. (b) Efeito do fator “Caracterização” sobre a variação do coeficiente de correlação.

6.3.5.4 Considerações finais sobre o efeito causado pela agregação de características

A associação de características locais da imagem com as divergências *Logistic loss* e *Generalized I-divergence* proporcionam um melhor desempenho no registro das imagens da amostra, quando comparadas a combinação tradicional: *Squared loss* e cor, utilizada em grande parte dos métodos de registro. Sendo assim a generalização propiciada pelo funcional proposto mostra-se bastante relevante para o processo de registro.

Dos experimentos apresentados constata-se ainda que tanto o fator “Divergência” quanto o fator “Caracterização” influenciam no desempenho do registro de imagens. A Tabela 24 apresenta os melhores resultados utilizando as divergências *Logistic loss* e *Generalized I-divergence*, obtidos nos experimentos III a VII, utilizando outras informações além da cor.

6.3.6 Comparação de métodos

Com o intuito de avaliar o desempenho do método proposto, propõe-se a comparação com os resultados obtidos utilizando a função “imregister” disponível no *Image Processing toolbox* do *software Matlab*. A função “imregister” utiliza um processo iterativo para registrar imagens e tem como parâmetros a **métrica** que define a similaridade entre as imagens e o **otimizador** que define a metodologia para minimizar ou maximizar a

Tabela 24 – Coeficientes de correlação dos melhores resultados obtidos utilizando as divergências *Logistic loss* e *Generalized I-divergence* associadas a outras informações além da cor.

	<i>Squared loss</i> + cor	<i>Logistic loss</i> + outra	GID+ outra
Prostata	68,1%	86,8%	85,1%
Coração	36,6%	58,3%	57,2%
Abdômen (corte 1)	65,5%	77,0%	77,7%
Abdômen (corte 2)	79,7%	83,3%	80,0%

similaridade. Como otimizadores são disponibilizados o método do gradiente descendente e o método *One Plus One Evolutiva*, estratégia que simula a evolução biológica de um conjunto de amostras no espaço de busca. Para a escolha da métrica tem-se: *Means Square* dada pela soma das diferenças ao quadrado entre os valores de intensidade das imagens, e *Mutual Information* baseada no conceito da teoria da informação.

6.3.6.1 Experimento VIII

A Tabela 25 apresenta as medidas de avaliação CC e RMSE para registro das imagens analisadas nos experimentos I,II e III utilizando a função “imregister”. O objetivo deste experimento é testar a eficácia desta no registro das imagens sintéticas e médicas apresentadas nestes experimentos.

Tabela 25 – Avaliação registro de diferentes imagens (sintéticas e médicas) utilizando a função “imregister”.

	<i>CC_{inicial}</i>	CC	RMSE	Variação CC
Galinha	0,8457	0,9914	0,0002	13,4%
Retângulo	0,5856	0,1177	0,0021	-
Próstata	0,3317	0,7200	0,0003	58,1%
Coração	0,8900	0,9052	0,0005	13,7%
Peito	0,9419	0,9485	0,0007	53,0%
Abdômen (corte 1)	0,7421	0,7991	0,0006	22,1%
Abdômen (corte 2)	0,7514	0,8284	0,0005	10,2%

A função “imregister” foi aplicada considerando a métrica *Mutual Information* e o otimizador *One Plus One Evolutiva* e 3000.

Destes resultados destaca-se a ineficácia do registro do retângulo apresentado na Figura 44, enquanto que, o funcional proposto mostrou-se bastante eficaz ao realizar este registro, obtendo um coeficiente de correlação igual a 0,9875 conforme apresentado na Tabela 10.

Os resultados nos registros das outras imagens foram menos eficazes do que os alcançados utilizando o funcional proposto. Este comportamento pode ser resultado do número de iterações definido.

6.3.6.2 Experimento IX

A Figura 58 apresenta uma comparação dos registro das imagens de dois corte abdominais considerando as caracterização “cor” e “cor e entropia”. Em experimentos anteriores verificou-se que a associação de características “cor e entropia” com a divergência *Logistic loss* proporcionou um melhor desempenho no registro quando comparados aos resultados obtidos considerando as caracterizações isoladas e a distância *Squared loss*. Neste experimento pretende-se, além da comparação com o desempenho da função “imregister”, confirmar esta hipótese.

O método funcional apresentado foi aplicado combinando as caracterização “cor” e “cor e entropia” com as divergências *Squared loss* e *Logistic loss*. Para cada uma das quatro configurações obtidas foram realizadas 2500 iterações com passo $\Delta t = 0,05$ e $\lambda = 15$. De forma análoga ao Experimento II desta seção a função “imregister” foi executada com os parâmetros: optimizer.GrowthFactor=1.05, optimizer.Epsilon= 1.5e-06,optimizer.InitialRadius=8.9286e-04 e optimizer.MaximumIterations= 3000. O coeficiente de correlação inicial entre as imagens S e T, $CC(S, T)$ é igual a 0,7514. A Tabela 26 apresenta uma comparação dos coeficientes de correlação entre a imagem T e as imagens finais resultantes dos registros representados na Figura 58. O registro aplicando a função “imregister”, quando comparado com os outros métodos não foi satisfatório.

Tabela 26 – Comparação do desempenho dos métodos de registro propostos com o método disponibilizado pelo *software Matlab* - Cortes Abdominais

	CC	Variação CC
<i>Mutual Information e One Plus One Evolutiva</i>	0,9163	21,95%
<i>Logistic loss - “cor”</i>	0,9632	28,19%
<i>Logistic loss- “cor e entropia”</i>	0,9669	28,68%
<i>Squared loss - “cor”</i>	0,9445	25,70%
<i>Squared loss - “cor e entropia”</i>	0,9535	26,90%

Na Figura 59 pode-se verificar a evolução do coeficiente de correlação considerando 2500 iterações. Observa-se que embora o registro realizado utilizando a função “imregister”, disponível no software Matlab, proporcione um aumento do coeficiente de correlação de aproximadamente 22%, o funcional proposto é mais eficaz. Destaca-se o desempenho obtido associando as características cor e entropia com a divergência *Logistic loss*.

Para a caracterização “cor”, a divergência *Logistic loss* apresentou CC 5,2% maior que o CC obtido no registro utilizando a função *imregister*. A diferença com relação ao CC obtido com a combinação “cor” e divergência *Squared loss*, utilizada na maioria dos métodos de registro, é de 1,4%.

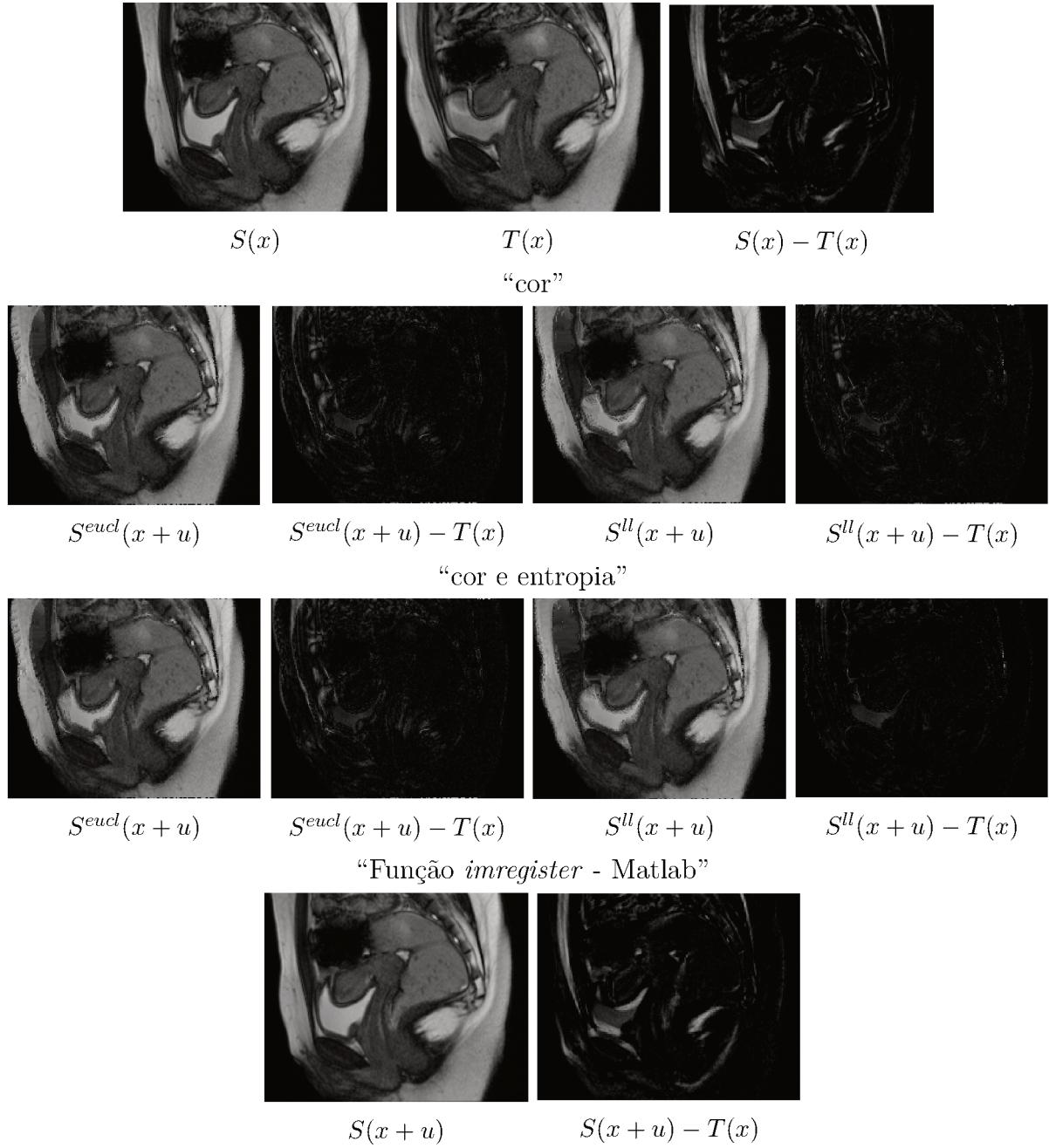


Figura 58 – Comparação de resultados: Linha 1 - Source e Target e a diferença inicial $S(x)-T(x)$ linha 2 - resultados e diferenças finais $S(x + u) - T(x)$ obtidos pelo funcional proposto associando a caracterização “cor” com as divergências *Squared loss* e *Logistic loss*, linha 3 - resultados e diferenças finais $S(x + u) - T(x)$ obtidos pelo funcional proposto associando a caracterização “cor e entropia” com as divergências *Squared loss* e *Logistic loss* e linha 4 - resultado e diferença final obtido pela função “*imregister*” do Matlab

6.3.6.3 Experimento X

Neste experimento aplicou-se o modelo de registro proposto na equação (64) para o par de imagens médicas S e T apresentadas na primeira linha da Figura 60 considerando

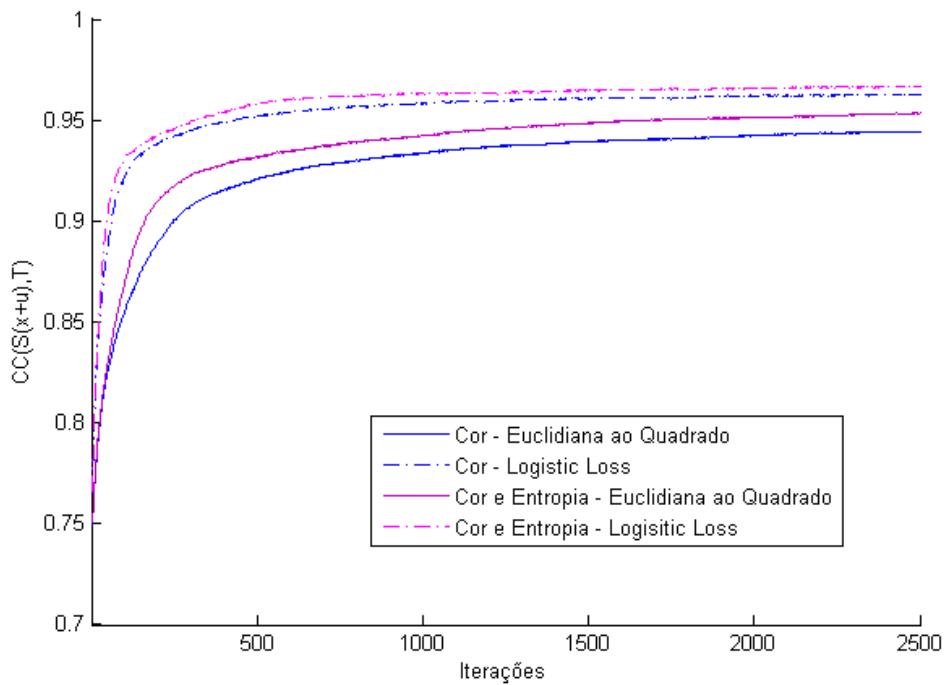


Figura 59 – Coeficiente de correlação x Iterações.

a caracterização “cor e entropia” e a divergência *Logistic loss*. Foram realizadas 100 iterações com passo $\Delta t = 0,05$ e $\lambda = 50$. O coeficiente de correlação inicial entre as imagens S e T, $CC(S, T)$ é igual a 0,5850.

A segunda e terceira linhas da Figura 60 apresentam, respectivamente, os resultados obtidos com o funcional proposto e com a função *imregister*. Os coeficientes de correlação final e $CC_{relative}$ resultantes destes registros são apresentados na Tabela 27 e demonstram o bom desempenho do método com a divergência *Logistic loss*.

Tabela 27 – Comparação do desempenho do método de registro proposto com o disponibilizado pelo *software Matlab - Cérebro*.

	CC	$CC_{relative}$
<i>imregister</i>	0.83	61%
<i>Logistic loss</i>	0.86	69%

A diferença entre os $CC_{relative}$ é de 8%, considerando apenas 100 iterações para a aplicação do método de registro proposto.

6.3.6.4 Experimento XI

As imagens de ressonância magnética (MRI) de um joelho apresentadas na Figura 61 são utilizadas em um exemplo no Matlab cujo objetivo é demonstrar a aplicação da função *imregister* no registro multimodal de imagens.

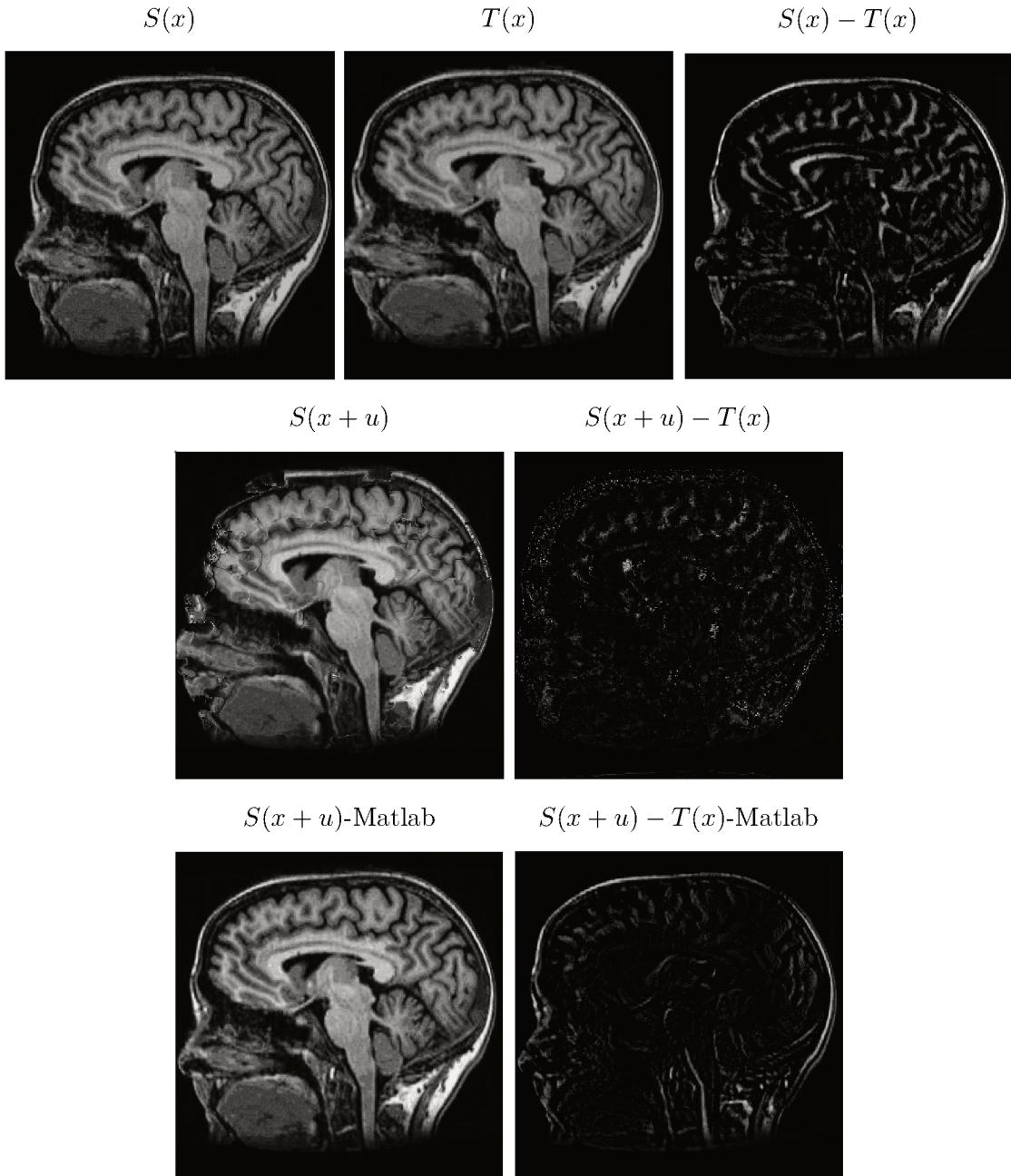


Figura 60 – Comparaçāo de resultados: Linha 1 - Source e Target e a diferença inicial $S(x)-T(x)$ linha 2 - resultado obtido pelo método e diferença final $S(x + u) - T(x)$, linha 3 - resultado obtido pela função “imregister” do Matlab;

Os dois corte sagitais S e T , ilustrados na primeira linha, foram adquiridos ao mesmo tempo e têm coeficiente de correlação inicial, $CC(S, T)$, igual a 0,6092.

A imagem $S(x + u)$ obtida com o registro realizado seguindo os passos apresentados neste exemplo é apresentada na quarta linha desta Figura. A função *imregister* foi aplicada com os parâmetros: GrowthFactor=1,05, Epsilon = 1,5e-06, InitialRadius = 8,9286e-04, MaximumIterations = 3000 e a métrica *Mutual Information*. O coeficiente de correlação entre a imagem final $S(x + u)$ e T , $CC(S(x + u), T)$ é igual a 0,7538.

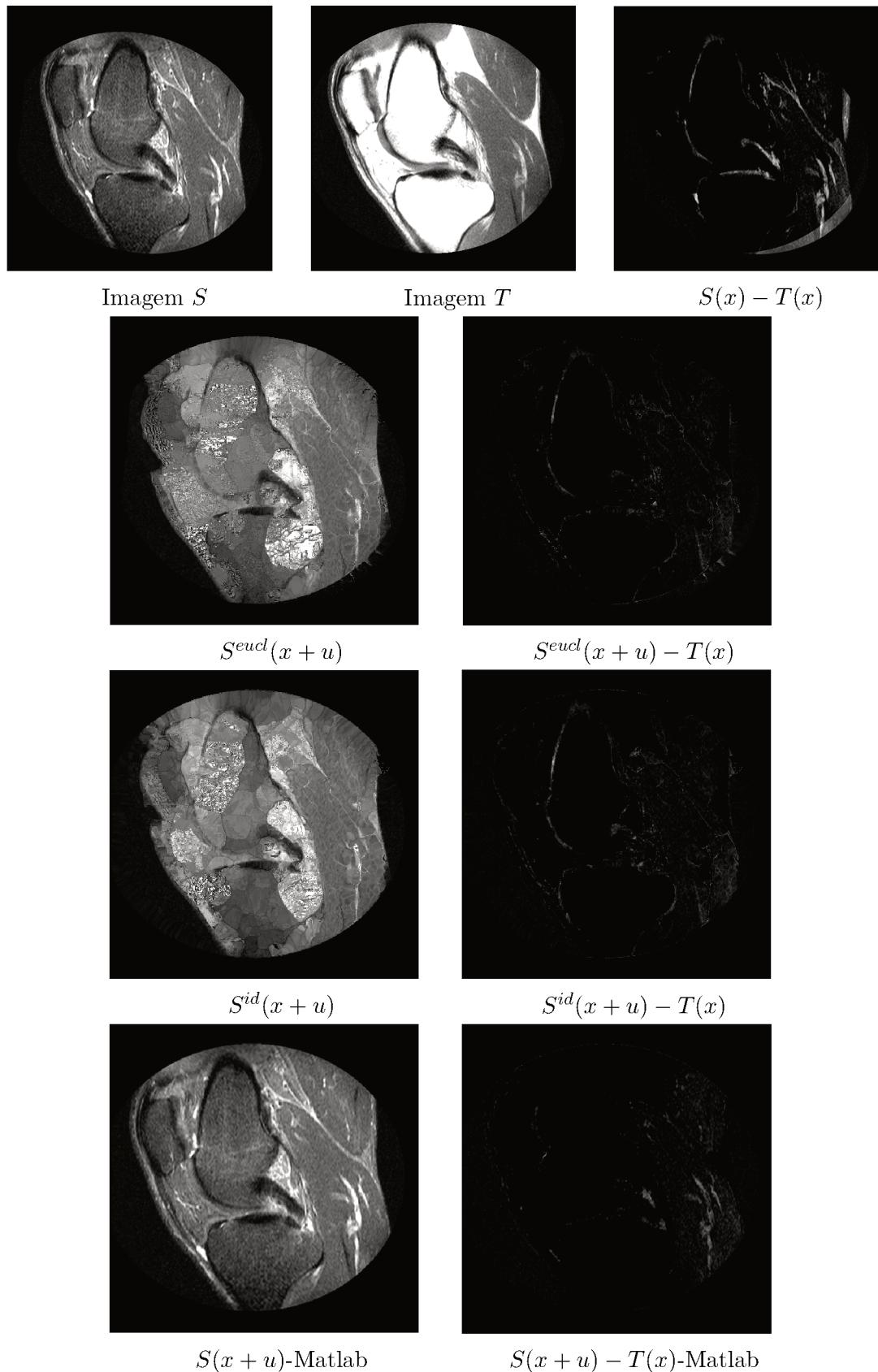


Figura 61 – Comparação de resultados: linha 1-Source e Target e a diferença inicial $S(x) - T(x)$, linha 2 - $S^{eucl}(x + u)$ e diferença final $S^{eucl}(x + u) - T(x)$, linha 3 - $S^{id}(x + u)$ e diferença final $S^{id}(x + u) - T(x)$ e linha 4 - resultado obtido pela função `imregister` do Matlab.

Com o intuito de avaliarmos o desempenho do funcional proposto no registro de imagens obtidas a partir de diferentes modalidades e scans, o método foi aplicado no registro destas imagens utilizando as divergência *Squared loss* e *Generalized I-divergence*.

Os resultados obtidos são apresentados nas segunda e terceira linhas da Figura, respectivamente. Para cada modelo foram realizadas 1000 iterações com passo $\Delta t = 0,05$ e $\lambda = 50$. Os coeficientes de correlação entre a imagem final $S(x + u)$ e T considerando a *Squared loss* e *Generalized I-divergence* são, respectivamente, $CC(S^{eucl}, T) = 0,8141$ e $CC(S^{id}, T) = 0,8107$.

Os registros proporcionados pelo nosso funcional têm coeficientes de correlação maiores quando comparados ao obtido seguindo o método apresentado no exemplo do Matlab. A diferença é de aproximadamente 6%. Este resultado demonstra a aplicabilidade do funcional proposto no registro multimodal.

6.3.6.5 Considerações finais sobre os experimentos de VIII a XI

O registro das imagens realizado com o funcional proposto é capaz de determinar de forma satisfatória o campo de deformação $u(x)$, utilizando diferentes divergência de Bregman, para o registro de imagens obtidas a partir de diferentes modalidades.

O funcional proposto possibilita a aplicação de configurações que tornam o registro mais eficaz quando comparado aos resultados alcançados mediante a aplicação da função *imregister*.

6.4 Considerações finais

Neste capítulo foi proposto um novo funcional que possibilita a aplicação das divergências de Bregman em problemas de registro de imagens. As equações de Euler-Lagrange e as de fluxo, associadas ao funcional proposto, foram deduzidas e utilizadas para minimizá-lo.

Apresentamos um meta-algoritmo que pode ser aplicado utilizando qualquer divergência de Bregman associada a diferentes caracterizações. Esta nova formulação variacional unifica vários métodos de registro possibilitando ao leitor determinar a melhor forma de construir e resolver o problema variacional de registro quando confrontados com uma nova aplicação.

Da análise dos resultados tem-se que o registro das imagem realizado com funcional proposto tem bom desempenho e é capaz de determinar de forma satisfatória o campo de deformação $u(x)$ para o registro de imagens obtidas a partir de diferentes scans ou modalidades.

Os experimentos mostram que a associação de características locais da imagem com as divergências *Logistic loss* e *Generalized I-divergence* proporcionam um melhor desempenho no registro das imagens da amostra, quando comparadas com a combinação tradicional:

Squared loss e cor, utilizada em grande parte dos métodos de registro. Sendo assim a generalização propiciada pelo funcional proposto mostra-se bastante relevante para o processo de registro.

É importante ressaltar que o método proposto apresenta algumas limitações. Em nosso modelo, o parâmetro de balanceamento λ é determinado experimentalmente, o que dificulta a sua aplicação. Além disso o valor deste parâmetro permanece inalterado (fixo) durante todo o processo de registro tornando o método menos preciso e necessitando de maior tempo para convergir. No próximo capítulo estenderemos este trabalho introduzindo um funcional com valor variável de λ .

CAPÍTULO 7

Divergências Total Bregman aplicadas no registro não rígido de imagens

A formulação variacional do problema de registro envolve a minimização de funcionais não convexos com relação a uma deformação desconhecida. Com o intuito de assegurar que o problema de minimização seja bem posto, em geral, o funcional é definido por um termo de regularização adicionado a uma medida de similaridade. O termo regularizador ou termo de suavização controla o campo de deformação desejado garantindo que este se mantenha suave durante o processo. Geralmente, a aplicação do método de registro exige a definição de um parâmetro, γ , que tem como papel balancear estes termos da equação. Apesar dos bons resultados obtidos, os métodos variacionais que adotam esta formulação são sensíveis à escolha deste parâmetro e para diferentes valores de λ tem-se diferentes resultados de registro. A escolha do termo de balanceamento é um grande desafio e em geral é realizada empiricamente. Esse desafio pode ser resolvido definindo um fator de ponderação auto-ajustável que torna a escolha do parâmetro de balanceamento mais flexível.

Os autores em (CHEN; YE, 2010) apresentaram um modelo no qual a razão entre os termos de similaridade e suavização é dada por $\gamma = \lambda/\sigma^2$ ao invés de simplesmente $\gamma = \lambda$, onde σ é o desvio padrão da imagem residual e λ uma constante. Embora a utilização deste fator de ponderação tenha gerado bons resultados, ainda há questões a considerar: a constante λ ainda deve ser determinada e o parâmetro de balanceamento é único para todos os pares de pontos das imagens a serem registradas. Computacionalmente isto aumenta o tempo necessário para a convergência.

Uma alternativa, proposta neste trabalho, para a solução desta questão é utilizar a divergência Total Bregman na definição do termo de similaridade. O fator conformal presente na definição da DTB torna esta divergência “adaptativa” na medida que considera a similaridade entre as imagens de forma ponderada.

Neste capítulo, propomos dois novos funcionais utilizando as divergências Total Bregman na definição do termo similaridade. No primeiro funcional, a imagem residual dada

pela divergência Total Bregman entre as imagens envolvidas no registro é utilizada como termo de similaridade. Em contraste com o que é apresentado em (CHEN; YE, 2010), este método contém, além de um parâmetro de balanceamento auto-ajustável, um fator de ponderação que varia em cada ponto da imagem. Esta capacidade surge naturalmente pois a divergência Total Bregman permite a minimização de uma soma ponderada das divergências entre os pares de pontos das imagens.

Em nossa segunda proposta, o novo funcional representa o problema de registro de imagens a partir das divergências Total Bregman utilizando como termo de similaridade a função log-verossimilhança negativa da imagem residual. Desta forma, tornamos o modelo menos sensível à escolha do parâmetro γ . Nesta nova proposta além da ponderação obtida com o fator conformal, o parâmetro de balanceamento é ajustado em função do desvio padrão da imagem residual. O processo inicia-se com um pequeno valor para γ e incrementa-o na medida que o erro de registro diminui.

Os funcionais propostos, utilizando diferentes escolhas de divergências, foram aplicados no registro de imagens médicas. Os resultados indicam que o uso das divergências Total Bregman podem propiciar maior eficácia e eficiência no registro de imagens.

7.1 Modelo de registro com Divergências Total Bregman

Sejam S e T duas imagens de um mesmo objeto definidas em Ω_S e $\Omega_T \subset \mathbb{R}^d$, respectivamente e $u : \Omega_S \rightarrow \Omega_T$ com $u(x) = (u_1(x), u_2(x), \dots, u_d(x))$, uma transformação vetorial, tal que a imagem modelo T e a imagem transformada $S(x + u(x))$, sejam similares considerando alguma divergência Total Bregman.

Propomos um novo funcional para representar o problema de registro de imagens a partir das divergências Total Bregman. Considerando que o cálculo do fator conformal depende do segundo argumento da DTB o novo funcional é definido utilizando a imagem transformada como segundo argumento do termo de similaridade dado pela divergência Total Bregman. Sendo assim propomos para o registro das imagens S e T a minimização do seguinte funcional:

$$F(u) = \|\nabla u\|^2 + \lambda tD_\varphi(T, S^u), \quad (78)$$

onde λ é uma constante, $tD_\varphi(x, y)$ é a divergência Total Bregman e $S(x + u) = S^u$.

Assim,

$$F(u) = \|\nabla u\|^2 + \lambda \rho_\varphi(S^u) d\varphi(T, S^u), \quad (79)$$

onde,

$$\rho_\varphi(S^u) = \frac{1}{\sqrt{1 + \|\nabla\varphi(S^u)\|^2}}.$$

O campo de deslocamento que minimiza o funcional definido em (79) é dado por:

$$-\Delta u + \lambda \frac{\partial}{\partial u} (\rho_\varphi(S^u) d_\varphi(T, S^u)) = 0. \quad (80)$$

onde

$$\begin{aligned} \frac{\partial}{\partial u_i} (\rho_\varphi(S^u) d_\varphi(T, S^u)) &= \langle \nabla \rho_\varphi(S^u), \nabla S^u \mathbf{e}_i \rangle d_\varphi(T, S^u) + \rho_\varphi(S^u) \frac{\partial d_\varphi(T, S^u)}{\partial u_i}, \\ \nabla \rho_\varphi(x) &= \frac{-\langle \nabla \varphi, \nabla(\nabla \varphi) \rangle}{\sqrt{(1 + \|\nabla \varphi(S^u)\|^2)^3}}. \end{aligned}$$

Da definição de divergência de Bregman, segue que:

$$d_\varphi(T, S^u) = \varphi(T) - \varphi(S^u - \langle T - S^u, \nabla \varphi(S^u) \rangle),$$

e portanto, as derivadas de d_φ com relação a u_i , são dadas por:

$$\frac{\partial d_\varphi(T, S^u)}{\partial u_i} = \langle S^u - T, \langle \nabla(\nabla \varphi(S^u)), \nabla S^u \mathbf{e}_i \rangle \rangle, \quad (81)$$

onde \mathbf{e}_i é o vetor da base canônica. As equações de Euler obtidas considerando as derivadas parciais das divergências Total Bregman são dadas por:

$$-\Delta u + \lambda [\langle \nabla \rho_\varphi(S^u), \nabla S^u \rangle d_\varphi(T, S^u) + \rho_\varphi(S^u) \langle S^u - T, \langle \nabla(\nabla \varphi(S^u)), \nabla S^u \rangle \rangle] = 0, \quad (82)$$

e sua equação de fluxo,

$$\frac{\partial u}{\partial t} = \Delta u - \lambda [\langle \nabla \rho_\varphi(S^u), \nabla S^u \rangle d_\varphi(T, S^u) + \rho_\varphi(S^u) \langle S^u - T, \langle \nabla(\nabla \varphi(S^u)), \nabla S^u \rangle \rangle]. \quad (83)$$

Para $S, T : \mathbb{R}^2 \rightarrow \mathbb{R}$ a função convexa φ é função de uma variável. Portanto $\nabla \varphi = \varphi'$ e $\nabla \nabla \varphi = \varphi''$. Assim, por exemplo, a equação de fluxo associada ao registro utilizando a divergência *Total Squared loss* é dadas por,

$$\frac{\partial u}{\partial t} = \Delta u - \lambda \left[\left\langle \frac{-4S^u}{\sqrt{(1 + 4(S^u)^2)^3}}, \nabla S^u \right\rangle (S^u - T)^2 + \left\langle \frac{2S^u}{\sqrt{1 + 4(S^u)^2}}, \nabla S^u \right\rangle (S^u - T) \right].$$

7.1.1 Detalhes da implementação

Para a implementação numérica as imagens, S e T , são representadas por matrizes de ordem $n \times m$, cujos valores são a intensidade em cada pixel. Seja u_{ij} , o valor do campo de deformação u associado ao pixel (i, j) com $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$. O campo de deformação é obtido iterativamente a partir da aplicação das equações de fluxo (83) nos tempos $t_k = k\Delta t$ com $k = 1, 2, \dots$, sendo Δt o tamanho do passo. A cada iteração k é

necessário a interpolação de $S(x + u)$, denotada por $\tilde{S}(x + u^k)$. Considerando $u(i, j, t_k) = u_{i,j}^k$, a forma discretizada da equação (83) é dada por:

$$\begin{aligned} u_{i,j}^{k+1} = u_{i,j}^k &+ \Delta t \left(\Delta u_{i,j}^k - \lambda \left[\langle \nabla \rho_\varphi \tilde{S}(x + u_{i,j}^k), \nabla \tilde{S}(x + u_{i,j}^k) \rangle d_\varphi(T(x), \tilde{S}(x + u_{i,j}^k)) + \right. \right. \\ &+ \rho_\varphi(\tilde{S}(x + u_{i,j}^k)) \left\langle \tilde{S}(x + u_{i,j}^k) - T(x), \langle \nabla(\nabla \varphi(\tilde{S}(x + u_{i,j}^k))), \nabla \tilde{S}(x + u_{i,j}^k) \rangle \right\rangle \left. \right] \right). \end{aligned} \quad (84)$$

7.2 Fator conformal

A DTB pode ser vista como a média ponderada da DB entre os pontos das imagens, com o peso inversamente proporcional à magnitude do gradiente da função convexa φ . Devido a convexidade, φ tem função gradiente crescente. Desta forma, o parâmetro de balanceamento terá menos influência no termo de similaridade quando os valores dos pontos da imagem T forem altos.

Considere por exemplo as imagens S e T apresentadas na Figura 62. As diferenças entre os níveis de cinza dos pontos da imagem S e seus correspondentes em T , obtidas com as divergência *Squared loss* são todas iguais a 0,16. Entretanto se aplicarmos a divergência *Total Squared loss* a divergências entre os pontos tem valores diferentes. As imagens tD^{sl} e d^{sl} apresentam, respectivamente, as divergências *Total Squared loss* e *Squared loss* entre S e T .



Figura 62 – Influência do fator conformal.

Observe que $tD^{sl}(0, 1; 0, 5) = tD^{sl}(0, 9; 0, 5)$ enquanto que $tD^{sl}(0, 5; 0, 1) > tD^{sl}(0, 5; 0, 9)$. Se a divergência *Squared loss* ordinária, ou seja DB, for utilizada, o resultado é o mesmo para todos os pares.

Assim, no funcional proposto em (82) o fator conformal $\rho_\varphi(S(x + u))$ é um termo de ponderação associado à divergência de cada par de pontos, além disso depende do campo de deslocamento u , o qual se modifica durante o processo de registro. Desta forma tem-se que o termo de balanceamento é ponderado para cada ponto $x = (i, j)$ e em cada iteração do método. A Figura 63 ilustra o fator conformal em diferentes iterações de um processo de registro. Observe que os valores variam em cada posição da imagem.

O fator conformal é um fator de ponderação auto-ajustável que torna o termo de similaridade adaptativo e acelera a convergência do método.

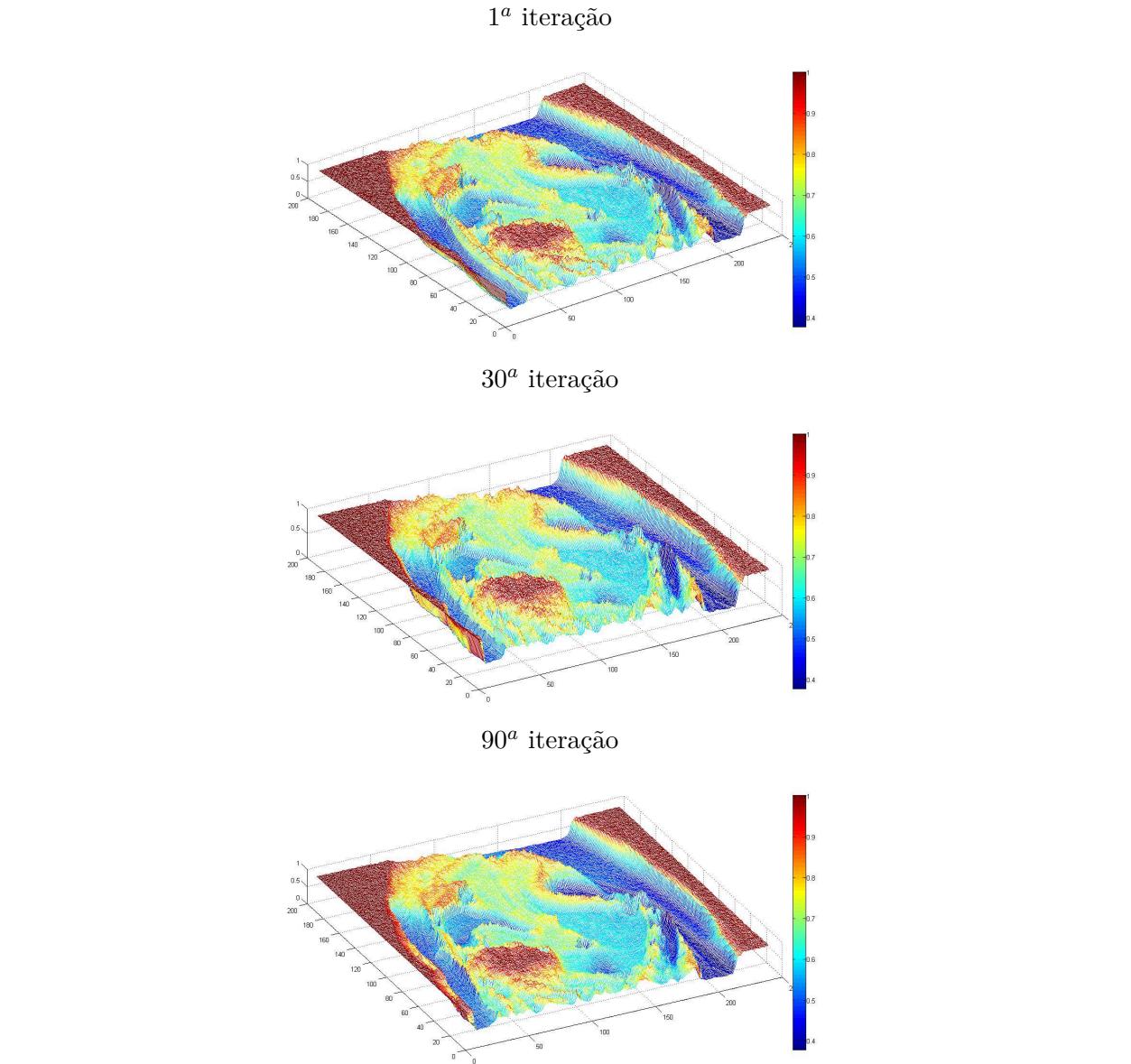


Figura 63 – Imagens do fator conformal obtidas em diferentes iterações do processo de registro.

7.3 Modelo de registro com Divergências Total Bregman e parâmetro de balanceamento auto-ajustável

No método de registro apresentado na seção anterior o termo de balanceamento λ tem seu efeito ponderado em cada pixel (x, y) devido ao fator conformal presente na definição da DTB. Entretanto a aplicação do método ainda exige a definição de um valor ótimo, a qual é realizada empiricamente. Esta característica dificulta a aplicação do método de registro proposto.

Com o intuito de tornar a escolha do parâmetro de balanceamento mais flexível propomos uma modificação no funcional proposto em (78). Nesta nova proposta o termo de similaridade é dado pela função log-verossimilhança negativa da imagem residual. Seja

$W(x)$, a intensidade de cada ponto da imagem residual entre $S(x + u)$ e $T(x)$ dada por:

$$W(x) = \sqrt{tD_\varphi(T(x), S(x + u))}; x \in \Omega = \Omega_T. \quad (85)$$

Considerando a intensidade dos pontos de W como amostras independentes de uma distribuição Gaussiana com média zero e variância σ^2 tem-se a seguinte função de verossimilhança:

$$\mathcal{L}(\sigma | W(x), x \in \Omega) = \prod_{x \in \Omega} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-|W|^2/2\sigma^2} \right).$$

Assim a função log-verossimilhança negativa é dada por:

$$-\int_{x \in \Omega} \left(-\log \sqrt{2\pi}\sigma - \frac{|W|^2}{2\sigma^2} \right) dx = \frac{tD_\varphi(T, S^u)}{2\sigma^2} + |\Omega| \log \sqrt{2\pi}\sigma.$$

O ponto que maximiza a verossimilhança é o mesmo que minimiza a função log-verossimilhança negativa. Sendo assim, omitindo a constante $\Omega \log \sqrt{2\pi}$, propomos para o registro das imagens S e T a minimização do seguinte funcional:

$$F(u, \sigma) = \|\nabla u\|^2 + \lambda \left(\frac{tD_\varphi(T, S^u)}{2\sigma^2} + |\Omega| \log \sigma \right). \quad (86)$$

onde λ é uma constante, $tD_\varphi(x, y)$ é a divergência Total Bregman e $S^u = S(x + u)$.

Assim,

$$F(u, \sigma) = \|\nabla u\|^2 + \frac{\lambda}{\sigma^2} \left(\rho_\varphi(S^u) d_\varphi(T, S^u) + |\Omega| \log \sigma \right) \quad (87)$$

onde,

$$\rho_\varphi(S^u) = \frac{1}{\sqrt{1 + \|\nabla \varphi(S^u)\|^2}}$$

O campo de deslocamento que minimiza o funcional definido em (86) é dado por:

$$\begin{cases} -\operatorname{div} \left(\frac{\partial \|\nabla u\|^2}{\partial \nabla u} \right) + \frac{\lambda}{\sigma^2} \left(\frac{\partial}{\partial u} (\rho_\varphi(S^u) d_\varphi(T, S^u)) \right) = 0, \\ \sigma^2 = \frac{tD_\varphi(T, S^u)}{|\Omega|} = \frac{W^2}{|\Omega|}. \end{cases} \quad (88)$$

Considerando as derivadas parciais das divergências Total Bregman dadas em (81) e W a imagem residual definida em (85), as equações de Euler são:

$$\begin{cases} -\Delta u + \frac{\lambda}{\sigma^2} [\langle \nabla \rho_\varphi(S^u), \nabla S^u \rangle d_\varphi(T, S^u) + \rho_\varphi(S^u) \langle S^u - T, \langle \nabla(\nabla \varphi(S^u)), \nabla S^u \rangle \rangle] = 0 \\ \sigma = \frac{W}{|\Omega|^{1/2}}. \end{cases} \quad (89)$$

Temos assim as seguintes equações de fluxo,

$$\frac{\partial u}{\partial t} = \Delta u - \frac{\lambda}{\sigma^2} [\langle \nabla \rho_\varphi(S^u), \nabla S^u \rangle d_\varphi(T, S^u) + \rho_\varphi(S^u) \langle S^u - T, \langle \nabla(\nabla \varphi(S^u)), \nabla S^u \rangle \rangle]. \quad (90)$$

O nosso funcional permite aplicação do método para diferentes DTB. Por exemplo, as equações de Euler Lagrange associadas ao registro com parâmetro de balanceamento $\gamma = \lambda/\sigma^2$ ajustável utilizando a divergência *Total Squared loss* são,

$$\begin{cases} \Delta u - \frac{\lambda}{\sigma^2} \left[\left\langle \frac{-4S^u}{\sqrt{(1+4(S^u)^2)^3}}, \nabla S^u \right\rangle (S^u - T)^2 + \left\langle \frac{2S^u}{\sqrt{1+4(S^u)^2}}, \nabla S^u \right\rangle (S^u - T) \right] = 0. \\ \sigma = \frac{|S^u - T|^2}{\sqrt{1+4(S^u)^2} |\Omega|}. \end{cases}$$

7.4 Detalhes da implementação

A divergência Total Bregman entre as imagens, S e T , é obtida pelo somatório das divergências entre cada pixel, ou seja,

$$tD_\varphi(S, T) = \sum_{j=1}^m \sum_{i=1}^n tD_\varphi(S(i, j), T(i, j)),$$

ou seja,

$$tD_\varphi(S, T) = \sum_{j=1}^m \sum_{i=1}^n \rho_\varphi(T(i, j)) d_\varphi(S(i, j), T(i, j)) = \sum_{j=1}^m \sum_{i=1}^n \frac{d_\varphi(S(i, j), T(i, j))}{\sqrt{1 + \|\nabla \varphi(T(i, j))\|^2}}$$

onde ρ_φ é o fator conformal, d_φ é a divergência de Bregman e $\nabla \varphi$ o vetor gradiente de φ .

Sendo u_{ij} , o valor do campo de deformação u associado ao pixel (i, j) com $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, m$, Δt o tamanho do passo com $t_k = k\Delta t$ para $k = 1, 2, \dots$ e considerando $u(i, j, t_k) = u_{i,j}^k$, a forma discretizada do sistema de equações definido em (90) é:

$$\begin{aligned} u_{i,j}^{k+1} = u_{i,j}^k &+ \Delta t \left[\Delta u_{i,j}^k - \frac{\lambda}{(\sigma^k)^2} \left(\langle \nabla \rho_\varphi(\tilde{S}(x + u_{i,j}^k)), \nabla \tilde{S}(x + u_{i,j}^k) \rangle d_\varphi(T(x), \tilde{S}(x + u_{i,j}^k)) + \right. \right. \\ &\left. \left. + \rho_\varphi(\tilde{S}(x + u_{i,j}^k)) \langle \tilde{S}(x + u_{i,j}^k) - T(x), \langle \nabla(\nabla \varphi(\tilde{S}(x + u_{i,j}^k))), \nabla \tilde{S}(x + u_{i,j}^k) \rangle \rangle \right) \right]. \end{aligned} \quad (91)$$

O valor de σ , em cada passo, segue da equação 89

$$\sigma^k = \sqrt{\frac{1}{|\Omega|} tD_\varphi(T(x), \tilde{S}(x + u^k))}. \quad (92)$$

$\tilde{S}(x + u^k)$ é a interpolação de $S(x + u)$ realizada em cada iteração k .

O modelo proposto torna a escolha do parâmetro de balanceamento γ mais flexível. Em cada iteração do método a razão entre o termo de similaridade e de suavização ao invés de $\gamma = \lambda$ é dado por $\gamma = \lambda/\sigma^2$. Como σ é otimizado a partir das equações de Euler temos que o peso do termo de similaridade é atualizado em cada iteração.

7.5 Experimentos e Análise de Resultados

Nesta seção apresentam-se experimentos realizados com o intuito de avaliar a eficácia dos funcionais proposto em (79) e (86). Os modelos de registro foram aplicado a imagens médicas e sintéticas, utilizando diferentes escolhas para a divergência Total Bregman com o intuito de: 1) avaliar a eficácia do registro com a DTB comparando-o com os métodos de registro definidos com a DB, 2) analisar o efeito que algumas DTB exercem no registro e 3) avaliar o desempenho do método considerando a aplicação com ajuste do valor de γ em função do desvio padrão da imagem residual conforme definido em (89).

7.5.1 Experimento I

O objetivo deste experimento é analisar o impacto do fator conformal no processo de registro. O modelo de registro proposto na Equação 82 foi aplicado ao par de imagens sintéticas ilustradas na Figura 64 utilizando a divergência Total Bregman: *Total Squared loss*.

Para avaliar o efeito do fator conformal, o mesmo par de imagens foi também registrado utilizando o método de registro apresentado no capítulo 6 com medida de similaridade definida pela divergência *Squared loss*. O desempenho dos dois métodos foi mensurado calculando o coeficiente de correlação e RMSE entre as imagens finais e imagem modelo, dados, respectivamente, em (23) e (24). Mediante a comparação destes resultados avaliou-se o efeito do fator conformal.



Figura 64 – Imagens S, T e Diferença Inicial.

A aplicação dos métodos de registro com a divergências *Total Squared loss* e *Squared loss* exige a definição do parâmetro de balanceamento λ . Para este experimento considerou-se o fator $\gamma = \lambda$ onde λ é constante e estabelecido empiricamente. Após a definição dos valores de λ adequados aos dois métodos, comparamos os resultado dos registros obtidos com as duas divergências e avaliamos o impacto do fator conformal no processo. Em seguida o modelo de registro com parâmetro γ auto-ajustável definido em 89 e 90, foi avaliado.

Os testes foram realizados considerando diversos valores para a constante λ e as divergências *Squared loss* e *Total Squared loss*.

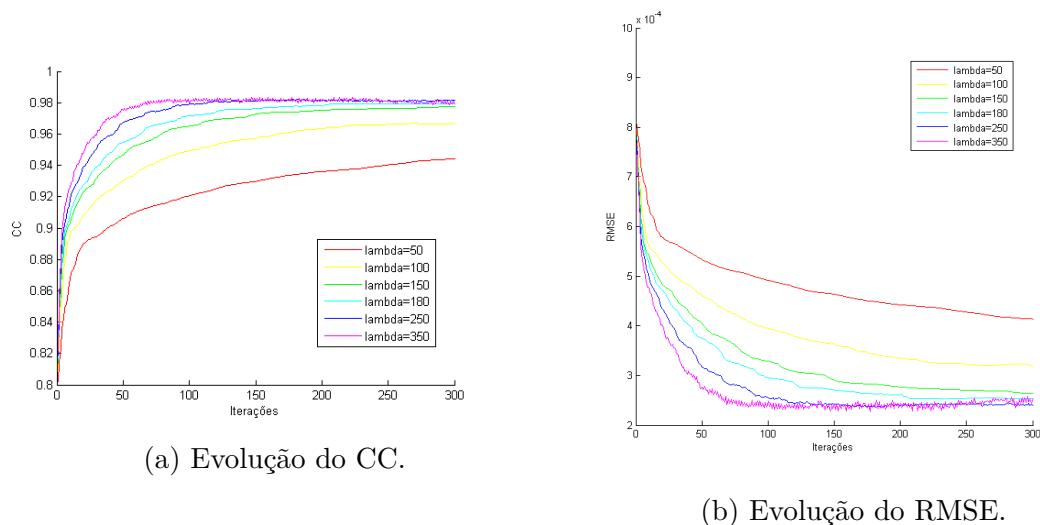


Figura 65 – Medidas de Avaliação do registro obtido com a divergência **Total Squared loss** para diferentes valores de λ

A variação do coeficiente de correlação e RMSE do método de registro considerando a divergência *Total Squared loss* e $\lambda = 50, 100, 150, 180, 250$ e 350 são apresentados, respectivamente, na Figura 65 (a) e (b). Observamos pelas figuras que para os diferentes valores

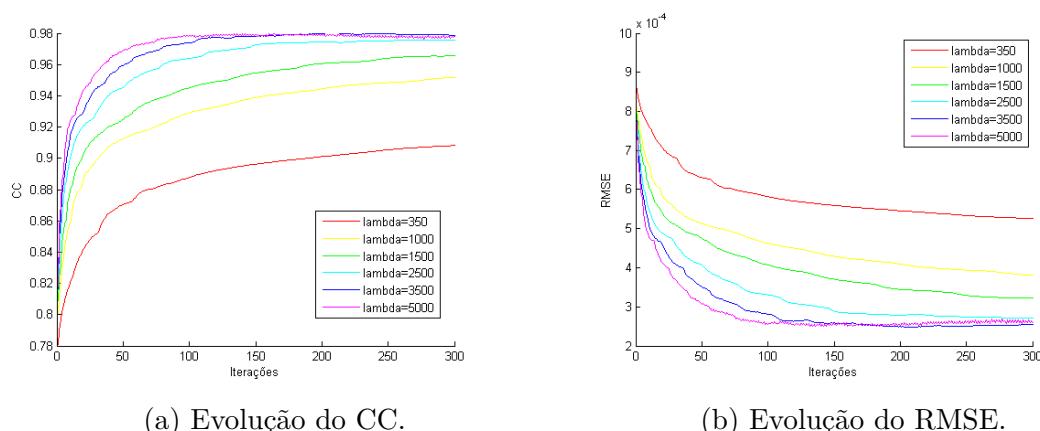


Figura 66 – Medidas de Avaliação do registro obtido com a divergência **Squared loss** para diferentes valores de lambda.

de λ o método de registro com a *Total Squared loss* apresenta comportamento similar e é convergente. A diferença apresenta-se maior nas 200 primeiras iterações.

Este mesmo comportamento pode ser observado no registro com a *Squared loss* com diferentes valores para o fator de balanceamento. As Figuras 66 (a) e (b) apresentam os resultados do CC e RMSE para o registro considerando $\lambda = 350, 1000, 1500, 2500, 3500$ e 5000 .

Temos que para os valores mais altos de λ os métodos utilizando a DTB e DB, embora estejam convergindo, oscilam muito em torno do valor de convergência. A utilização de parâmetros maiores do que os apresentados, tanto para a *Squared loss* quanto para a *Total Squared loss*, torna o processo divergente. Enquanto que para valores menores, nos dois métodos a convergência é lenta tornando o processo menos eficiente.

A Figura 67 apresenta os valores do CC em função do tempo de processamento para as duas situações, com DTB e DB, utilizando diferentes valores para λ , considerando 300 iterações. Observamos nos dois métodos que o tempo de processamento, é praticamente mesmo.

Entretanto se compararmos o tempo gasto no registro com a DTB e com DB, observamos que o registro que utiliza DTB exige um maior tempo de processamento. Por outro lado este método alcança mais rápido o valor de convergência.

Na Figura 68 apresentamos uma comparação entre os métodos utilizando como parâmetro de balanceamento $\lambda = 250$ e $\lambda = 3500$, para o registro com as divergências *Total Squared loss* e *Squared loss*, respectivamente. A escolha desses valores justifica-se pelo fato de serem valores para os quais os métodos convergem mais rápido e são mais suaves. Observa-se, Figuras 65 (a) e 66 (a), que para valores maiores que os definidos acima tem-se muitas oscilações e se aumentados levam à divergência computacional. A Figura 69 ilustra esse comportamento utilizando a divergência *Squared loss* com $\lambda = 700$. Observe que o valor do CC após 50 iterações cai drasticamente.

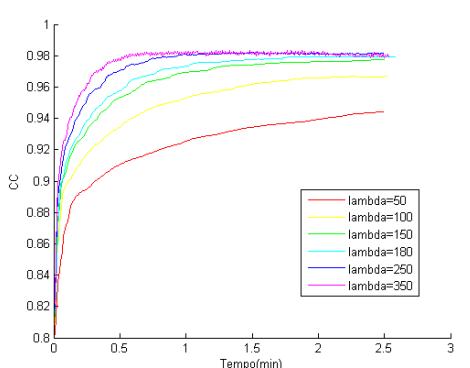
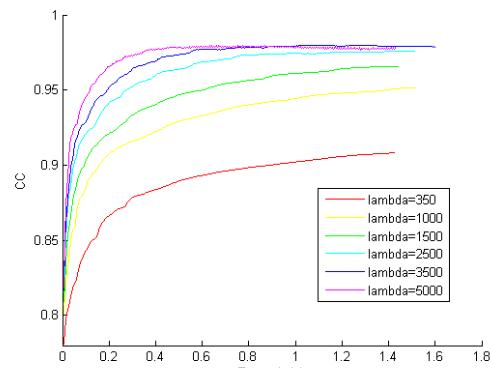
(a) *Total Squared loss*.(b) *Squared loss*.

Figura 67 – Valores do CC em função do tempo de processamento utilizando diferentes valores para λ .

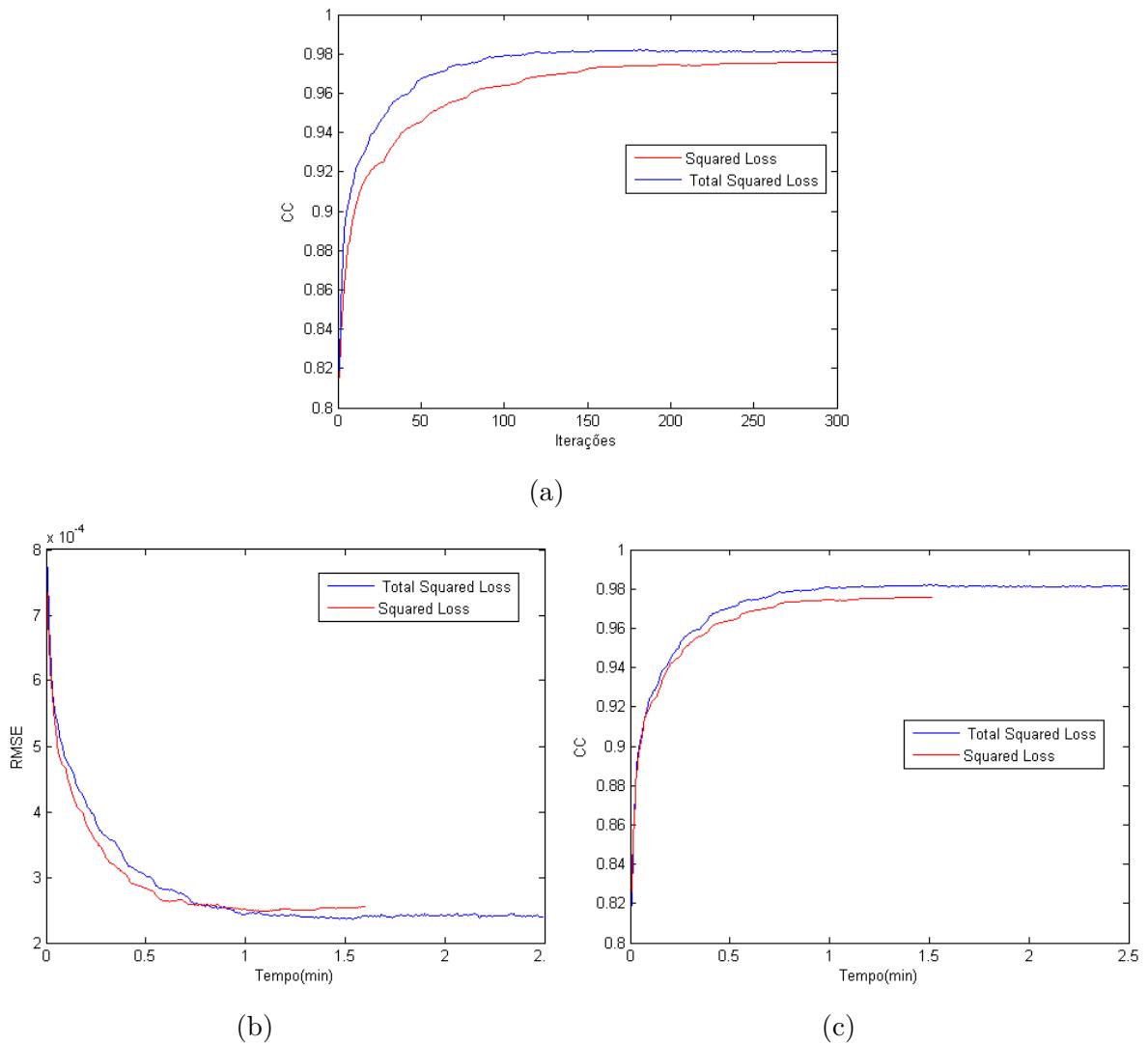


Figura 68 – Comparação entre os métodos registro com as divergências *Total Squared loss* e *Squared loss*, utilizando λ igual a 250 e 3500, respectivamente.

Note que o processo com a divergência *Squared loss* é mais eficiente, com tempo aproximadamente de 1,5 minutos (Figura 68(c)), porém o método com a *Total Squared loss* tem, durante todo o processo, coeficiente de correlação maior quando comparado com a *Squared loss*. Constatamos que o fator conformal, gerou um custo maior no tempo de processamento, visto que implica na realização de mais operações, mas proporcionou maior velocidade na convergência e tornou o método mais eficaz.

O fator conformal torna o parâmetro de balanceamento $\gamma = \lambda$ adaptativo durante o processo. Nas Figuras apresentamos a evolução do produto $\lambda\rho$ para os quatro pontos apresentados na Figura 71. Observe que para os pontos que no decorrer do processo tem o fator conformal alterado, na medida que a imagem S se aproxime de T o produto $\lambda\rho$ tende a se estabilizar.

A Figura 72 ilustra as imagens e os campos de deslocamento obtidos com os dois métodos de registro. Observa-se que o campo de deslocamento obtido com a divergência *Total*

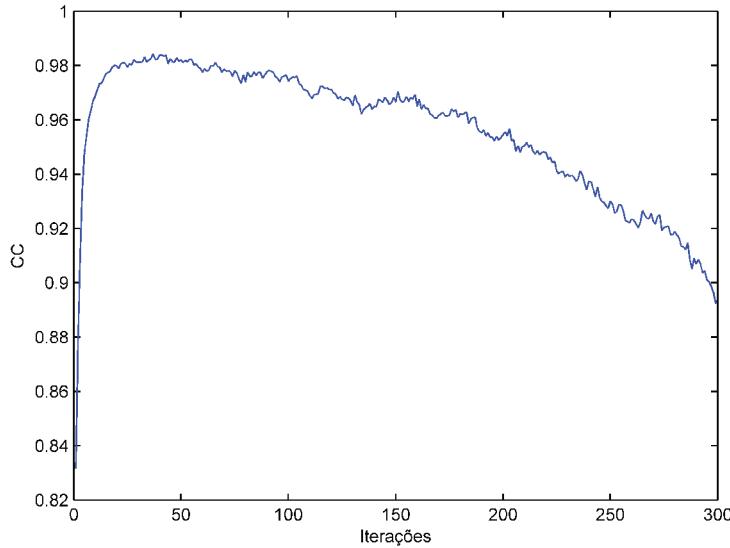


Figura 69 – Evolução do CC para o registro com divergência *Total Squared loss* e $\lambda = 700$.

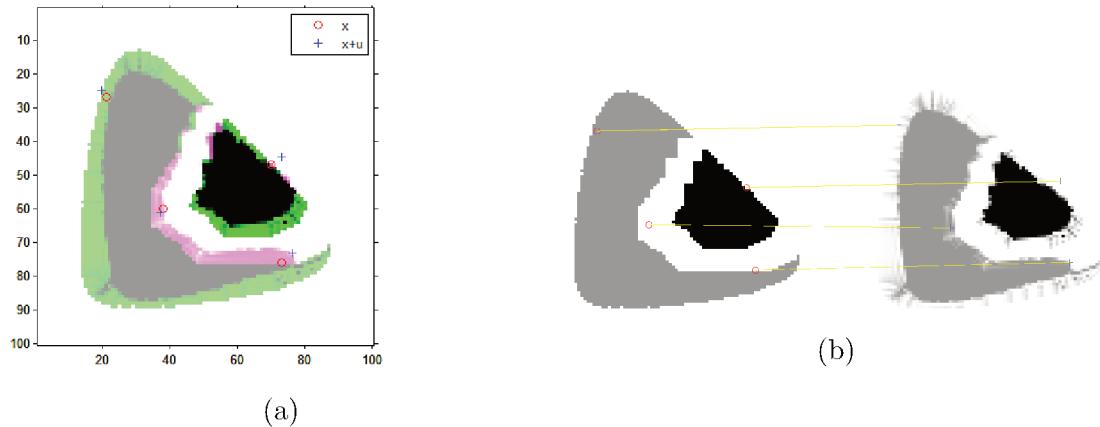


Figura 70 – Efeito do campo de deslocamento nos pontos $(73, 76)$, $(38, 60)$ $(70, 47)$ e $(21, 27)$.

Squared loss tem intensidade maior nas bordas do que o campo obtido com a divergência *Squared loss*.

7.5.2 Experimento II

Neste experimento o modelo de registro com parâmetro γ auto-ajustável definido nas Equações 89 e 90 foi aplicado ao par de imagens sintéticas ilustradas na Figura 64 utilizando a divergência *Total Squared loss*. A Figura 73 ilustra a evolução do CC para o método utilizando $\lambda = 0, 1$.

No experimento I este mesmo par de imagens foi registrado utilizando o funcional com parâmetro de balanceamento tendo como peso o fator conformal. Verificou-se experimentalmente que para o parâmetro *gamma*, fixo, dado por $\lambda = 250$, o método proposto convergia mais rápido de forma mais suave no registro das imagens.

O gráfico ilustrado na Figura 73 apresenta uma comparação entre os registros consi-

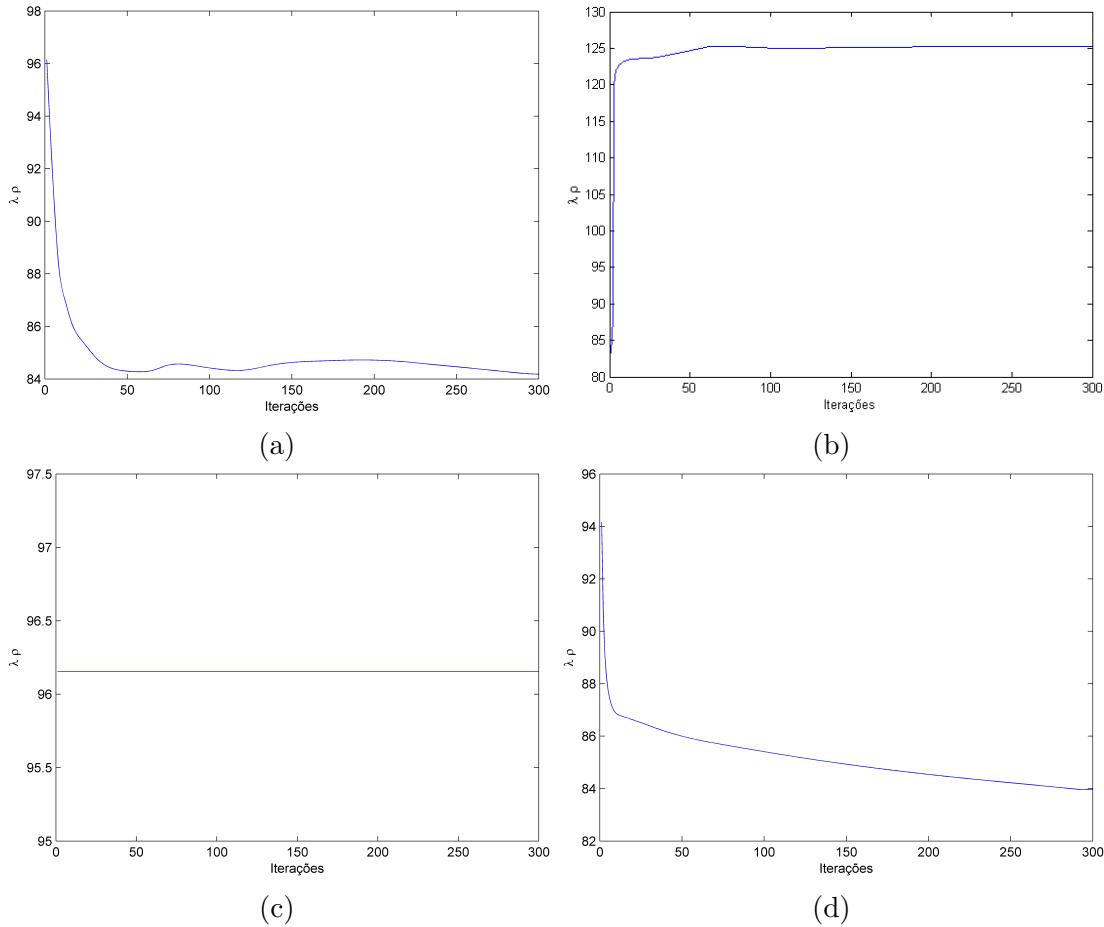


Figura 71 – Evolução do produto $\lambda\rho$ para os pontos: (a) (73, 76), (b) (38, 60), (c) (70, 47) e (d) (21, 27).

derando γ fixo igual a 250 e $\gamma = \lambda/\sigma$ com $\lambda = 0, 1$ para o processo com balanceamento variável.

A evolução do CC nos dois métodos é praticamente a mesma a partir das 70 primeiras iterações. Entretanto no método com parâmetro de平衡amento ajustado pelo desvio padrão da imagem residual a escolha desse parâmetro foi simplificada facilitando a sua aplicação. O método permite que no inicio do processo este parâmetro tenha valores menores, esta característica torna o registro mais eficiente e eficaz.

7.5.3 Experimento III

O método de registro com as divergências Total Bregman como termo de similaridade permite a utilização de qualquer divergência no processo de registro não rígido de imagens. Neste experimento avaliamos a influência das divergências *Total Squared loss*, *Total Logistic loss* e *Total I-divergence* na qualidade do registro das imagens ilustradas na Figura 46 do capítulo 6.

Aplicamos o modelo de registro com parâmetro auto-ajustável definido em (89) utilizando como critério de parada o tempo de processamento. Desta forma o parâmetro de

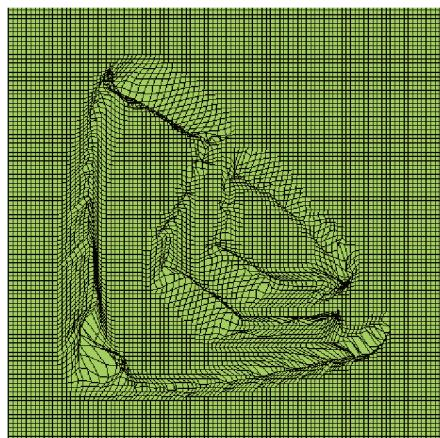
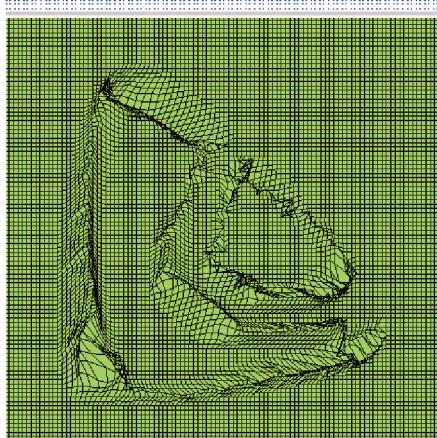
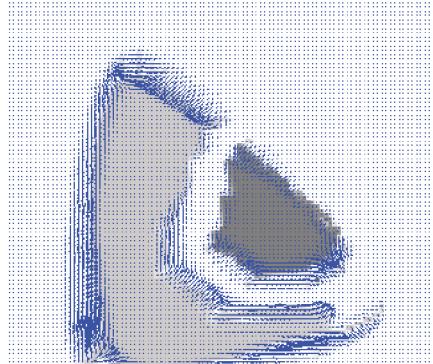
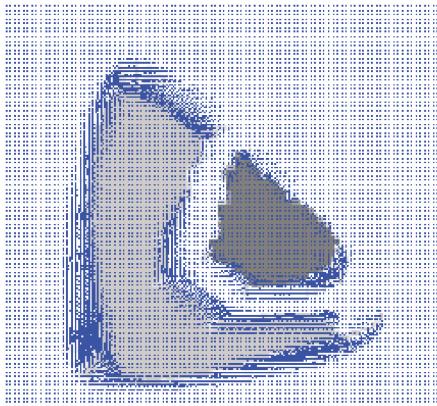
(a) *Total Squared loss*(b) *Squared loss*

Figura 72 – Em cada linha tem-se a imagem final S^u , a diferença final $S^u - T$, o campo de deslocamento u e a superfície de deslocamento, obtidos considerando o registro com as divergências (a) *Total Squared loss* e (b) *Squared loss*.

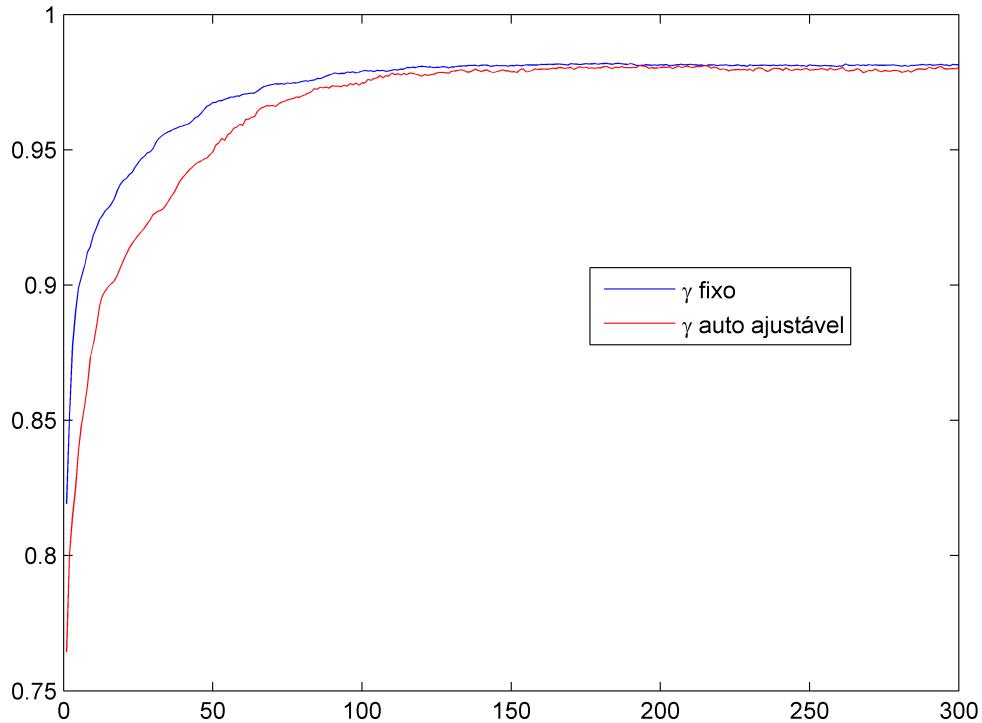


Figura 73 – Evolução do CC para os registros com γ fixo igual a 250 e γ variável com $\lambda = 0, 1$.

balanceamento foi ajustado em cada iteração e ponderado pelo fator conformal. Consideramos para todos os registros 5 minutos para o tempo máximo de processamento e lambda igual a 0,1. A Tabela 28 apresenta os coeficientes finais e relativos obtidos considerando os registros com cada uma das divergências.

Os resultados apresentados demonstram o bom desempenho do método considerando as diferentes divergências Total Bregman. A taxa de variação relativa do coeficiente de correlação aumentou consideravelmente para todas as opções de divergência, sendo que o método aplicado com a divergência *Total I-divergence*, em média, propiciou maior aumento relativo do CC. A divergência *Total Squared loss*, quando comparada às outras opções de divergência, obteve menor aumento do CC.

A Tabela 2 apresenta o número de iterações realizadas para cada um dos métodos. Observe que para a divergência *Total I-divergence* temos um número menor de iterações realizadas em todos os registro. Sendo também esta escolha a mais eficaz quando observamos o aumento relativo do CC (Tabela 28).

Por outro lado, embora tenhamos um bom desempenho do registro ao utilizar a divergência *Total Logistic loss*, similar à *Total I-divergence*, o número médio de iterações realizado no mesmo intervalo de tempo é superior à quantidade realizada pelas outras opções de divergências.

Desta forma, caso tivéssemos optado por um critério de parada dado pelo número de iterações a escolha da divergência *Total I-divergence*, também, tornaria o processo mais eficaz.

Tabela 28 – Medidas de avaliação obtidas nos registros das imagens utilizando o funcional com parâmetro $\gamma = \lambda/\sigma$ e as divergências *Total Squared loss*, *Total Logistic loss* e *Total I-divergence*.

	CC Inicial	CC Final			CC Relativo		
		Total SL	Total Log.loss	Total I-div.	Total SL	Total Log.loss	Total I-div.
Próstata	0,3317	0,6044	0,6408	0,8128	41%	46%	72%
Coração	0,8900	0,9736	0,9744	0,9780	71%	72%	76%
Peito	0,9419	0,9865	0,9912	0,9911	77%	85%	85%
Abdômen (Corte 1)	0,7421	0,9321	0,9440	0,9317	74%	78%	73%
Abdômen (Corte 2)	0,7514	0,8947	0,8446	0,9046	58%	37%	62%

Tabela 29 – Número de Iterações realizadas para t=5 min.

	Nº Iterações		
	Total SL	Total Log.loss	Total I-div.
Próstata	79	97	86
Coração	610	637	346
Chest	621	640	601
Abdômen (corte 1)	44	45	44
Abdômen (corte 2)	135	139	130
Média	298	312	241

A Figura 74 ilustra as imagens finais e os campos de deslocamento obtidos para o registro dos cinco pares de imagens com esta divergência.

7.5.3.1 Comparação entre os métodos definidos com a DB e DTB

O funcional utilizando as DB, definido na equação 57 do capítulo 6, foi aplicado às imagens: próstata, coração, peito, abdômen (corte 1) e abdômen (corte 2), considerando como critério de parada o mesmo intervalo de tempo de 5 minutos e o ajuste do parâmetro de balanceamento em cada iteração dado por $\gamma = \lambda/\sigma$. De modo análogo, tomamos inicialmente λ igual a 0,1. A Tabela 30 apresenta uma comparação entre os resultados obtidos utilizando os funcionais definidos com a DB e DTB.

Dos resultados constatamos superioridade do método de registro que utiliza as divergências Total Bregman como termo de similaridade. Em todos os registros, considerando as três divergências Total Bregman testadas, tem-se maior aumento do CC. Em média o registro utilizando as DTB proporcionaram um CC relativo três vezes maior quando comparado com as DB correspondentes.

A diferença de velocidade entre os métodos pode ser visualizada na Tabela 31. Os métodos com as divergências totais realizam aproximadamente 150 iterações a mais no tempo de 5 minutos. Este comportamento é devido aos termos acrescentados nas equações de Euler, oriundos das derivadas do fator conformal.

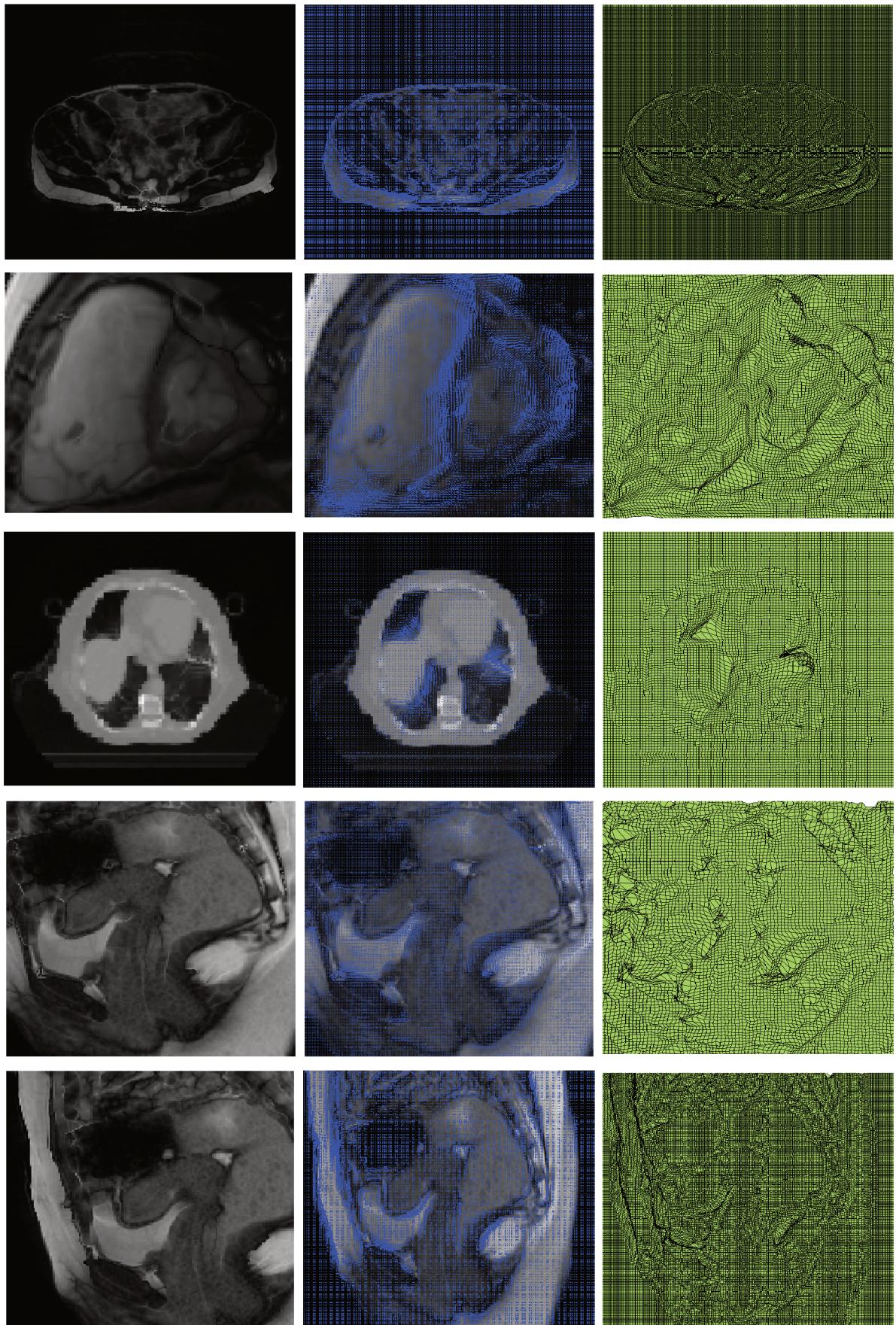


Figura 74 – Resultados finais para os pares de imagens: próstata, coração, peito, abdômen (corte1) e abdômen (corte2) utilizando a divergência *Total I -divergence*.

Tabela 30 – Comparação entre CC obtidos utilizando os funcionais definidos com a DB e DTB, considerando t=5 min e $\lambda = 0, 1$.

	SL	Total SL	Log.loss	CC_{rel}	GID	Total I.div
Próstata	13%	41%	20%	46%	18%	72%
Coração	30%	71%	44%	72%	44%	76%
Peito	19%	77%	23%	85%	21%	85%
Abdômen (Corte 1)	51%	60%	58%	67%	52%	60%
Abdômen (Corte 2)	45%	58%	51%	57%	50%	62%

Tabela 31 – Comparação entre o número de iterações obtidos utilizando os funcionais definidos com a DB e DTB, considerando t=5 min e $\lambda = 0, 1$

	SL	Total SL	Log.loss	Número de Iterações	GID	Total I-div.
Próstata	163	79	161	97	130	86
Coração	887	610	770	637	771	346
Peito	721	621	975	640	863	601
Abdômen (Corte 1)	336	144	299	145	304	144
Abdômen (Cortes2)	220	135	230	139	199	130
Média	465,4	318	487	331,6	453,4	261

7.6 Considerações finais

Neste capítulo propomos dois novos funcionais para representar o problema de registro de imagens a partir das divergências Total Bregman. Os método de registros com os novos funcionais permitem a minimização de uma soma ponderada das divergências entre os pixels das imagens tornando o processo “adaptativo”. Esta capacidade surge naturalmente devido ao fator conformal presente na definição da divergência Total Bregman. Desta forma, em nosso método o termo de balanceamento λ tem seu efeito ponderado em cada pixel (x, y) .

Além da ponderação obtida com o fator conformal, o parâmetro de balanceamento pode ser ainda ajustado em função do desvio padrão da imagem residual. Nesta nova proposta ao invés de adotarmos um parâmetro ótimo fixado no início do processo, iniciamos com um pequeno valor para λ e o incrementamos na medida que o erro de registro diminui. Assim a escolha desse parâmetro torna-se mais flexível facilitando a aplicação do método.

Os nossos funcionais permitem, ainda, a aplicação de qualquer DTB nos processos de registro propiciando um generalização relevante para o processo de registro.

Os funcionais com DTB foram testados em um conjunto de imagens médicas com o objetivo de avaliar a qualidade do registro com diferentes escolhas de divergências. Nossos resultados demonstraram que a DTB gera um custo no tempo de processamento,

mas proporciona maior velocidade na convergência e torna o método mais eficaz.

CAPÍTULO 8

Conclusão

No processamento e análise de imagens faz-se necessária a definição de medidas que sejam capazes de estabelecer a similaridade entre imagens, representadas, geralmente, por vetores, matrizes ou distribuições de probabilidades. Neste contexto, surgem aplicações nas quais a noção de similaridade mais adequada não pode ser definida por uma métrica. Contudo, a maior parte das estruturas e algoritmos, aplicados no processamento de imagens e visão computacional, são desenvolvidos considerando medidas de similaridade com propriedades de uma métrica.

Neste trabalho buscou-se superar essas limitações apresentando métodos de análise de imagens que utilizam as classes não métricas definidas pelas divergências de Bregman e Total Bregman.

As principais contribuições desta tese são a definição de três novos métodos que envolvem a minimização de funcionais que utilizam as divergências de Bregman ou total Bregman, uma nova estrutura de dados e um novo método de busca definidos com a DTB, e ainda, foram estabelecidos condições e tratamentos adequados para apoiar a pesquisa por similaridade dadas por estas divergências:

Método de segmentação hierárquica: utilizando o conceito de perda de informação de Bregman foi proposto um novo funcional para a segmentação de imagens. O algoritmo construído emprega um critério de fusão que minimiza a perda da informação de Bregman. Demonstrou-se que o método de segmentação hierárquica proposto generaliza o algoritmo de minimização da forma discreta do funcional de Mumford-Shah. Os resultados apresentados, obtidos utilizando as divergências *Generalizes I-Divergence*, *Itakura Saito* e a *Logistic loss*, evidenciam o bom desempenho do algoritmo.

Método de Registro com DB: apresentamos um novo funcional de energia que generaliza o método variacional empregado no registro de imagens não rígidas. O método de registro apresentado por Fisher e Modersitzki em (FISCHER; MODERSITZKI, 2004) segue como um caso particular de nossa abordagem. O funcional proposto

é definido considerando a divergência de Bregman como medida de similaridade e pode ser aplicado com divergências arbitrárias. Além disso, o modelo variacional aqui proposto permite a implementação simultânea de diferentes informações locais para o processo de registro. Desta forma, é possível adicionar informações espaciais ao método de registro variacional não rígido, permitindo quantificar de maneira mais eficiente as diferenças entre as imagens envolvidas no processo. O modelo apresentado é capaz de combinar as divergências Bregman com diferentes características, o que amplia as possibilidades de aplicação do novo método.

Método de Registro com DTB: foram propostos dois novos funcionais para representar o problema de registro de imagens a partir das divergências Total Bregman. As equações necessárias para a implementação computacional foram obtidas a partir da minimização dos funcionais propostos, utilizando as equações de Euler Lagrange e as equações de fluxo. Os métodos de registro com os novos funcionais permitem a minimização de uma soma ponderada das divergências entre os pixels das imagens tornando o processo “adaptativo”. Além disso o parâmetro de balanceamento pode também ser ajustado em função do desvio padrão da imagem residual. Da análise dos resultados obtidos observou-se que os funcionais propostos têm bom desempenho e são capazes de determinar de forma satisfatória o campo de deformação $u(x)$, necessário para o registro.

Árvore de bolas Total Bregman e algoritmo de busca: apresentamos uma nova estrutura de dados e um novo algoritmo para NNS usando a divergência Total Bregman (TBD). A estrutura e o método de pesquisa proposto permitem o desenvolvimento de meta-algoritmos, que são aplicáveis a qualquer DTB. Apresentamos um algoritmo clusterização e de re-ranking, nos quais a estrutura e método de busca são aplicados. As experiências aqui apresentadas demonstram o bom desempenho da nova estrutura nos métodos NNS, evidenciando, portanto, a importância da nova proposta.

Tratamentos Os tratamentos foram aplicados na recuperação de imagens por conteúdo e apresentaram bons resultados, excedendo os obtidos pelas distâncias convencionais. A aplicação dos tratamentos definidos considerando as características dos dados e os domínios de aplicação das divergências propiciam uma otimização dos métodos aplicados na análise de imagens.

8.1 Produtos gerados

O objeto de estudo gerou 4 artigos em conferências qualificadas (A1, A1, A2 e B1), e 3 artigos em periódicos: 1 em fase de 2^a revisão (qualis A1), 1 submetido (qualis A1), outro referente ao capítulo 7 em fase final de escrita e organização dos dados.

8.2 Trabalhos futuros

Considerando a diversidade do trabalho de pesquisa apresentado nesta tese, algumas extensões são possíveis:

Representação semântica do conteúdo: aplicar a estrutura de árvore de bolas total Bregman para a definição do conjunto de conceitos semânticos definidos pelos t-centros dos agrupamentos.

Segmentação: aplicar a divergência total Bregman para a segmentação de imagens.

Recuperação de imagens por representantes: utilizar os t-centros para otimizar o processo de recuperação de imagens.

Métodos de registro com DB: identificar caracterizações mais adequada para diferentes tipos de divergências de Bregman no registro de imagens.

Métodos de registro com DTB: propor um modelo variacional que permita a implementação simultânea de diferentes informações locais para o processo de registro com a DTB.

Parâmetro de balanceamento variável: propor outras formas de ajuste para o parâmetro de balanceamento dos modelos variacionais definidos com as divergências de Bregman e total Bregman.

Métodos de registro consistente: aplicar as divergências de Bregman e total Bregman na definição de funcionais que possam ser utilizados em métodos nos quais a imagem de origem e de destino sejam deformadas simultaneamente. As imagens deformadas são alinhadas de maneira que as duas transformações sejam inversamente consistentes. Este processo acelera o registro das imagens e o torna mais eficaz.

A Figura 75 ilustra juntamente com os principais conceitos, contribuições e publicações associadas, as extensões mencionadas acima, as quais são destacadas em vermelho.

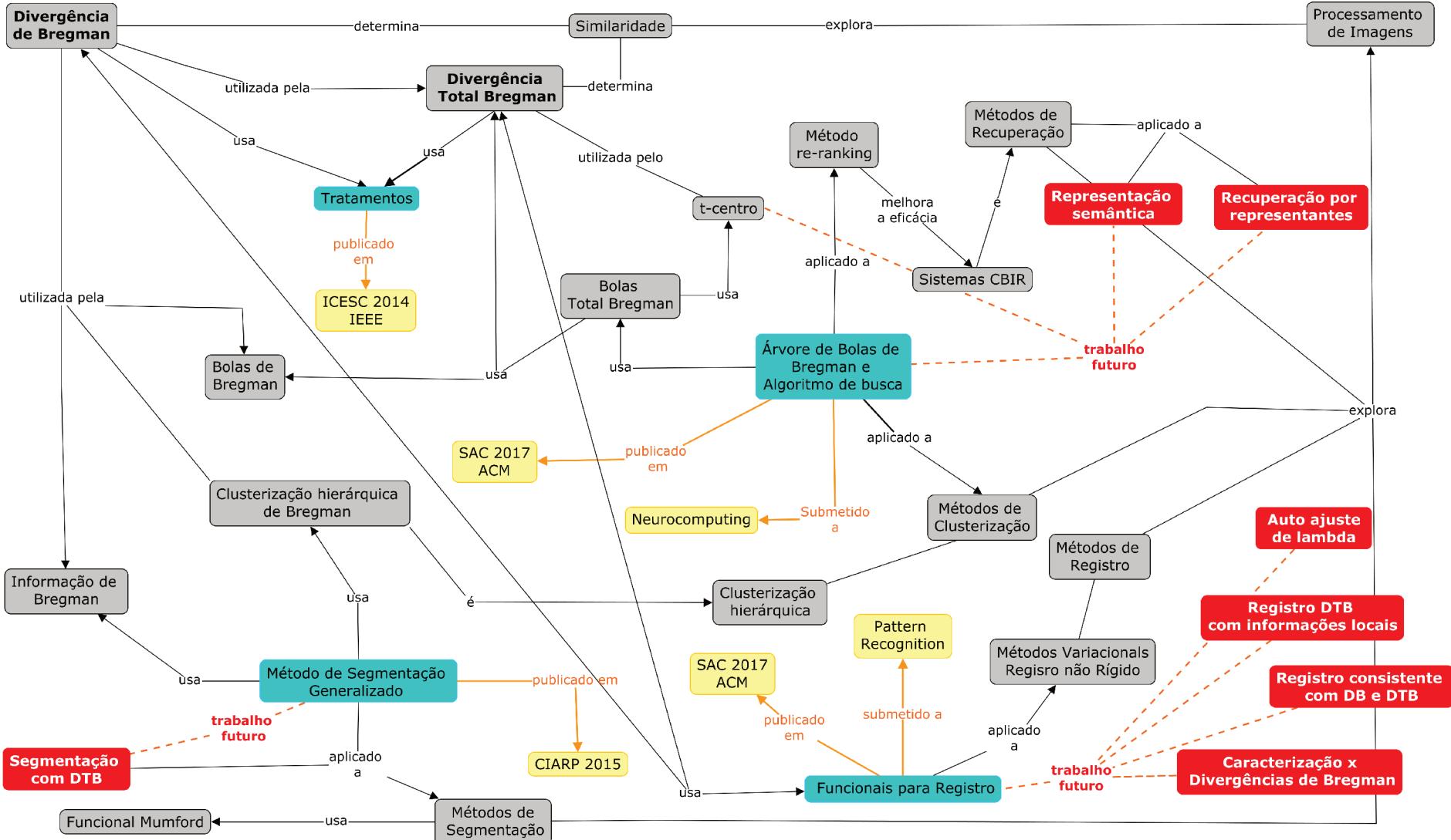


Figura 75 – Conceitos principais, contribuições, publicações e extensões.

Referências

ALDENDERFER, M. S.; BLASHFIELD, R. K. Cluster analysis. Beverly Hills Calif. Sage Publications 1984., 1984.

ALVAREZ, L. et al. Symmetrical dense optical flow estimation with occlusions detection. **International Journal of Computer Vision**, Springer, v. 75, n. 3, p. 371–385, 2007.

ANDONI, A.; INDYK, P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In: IEEE. **Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on.** [S.l.], 2006. p. 459–468.

BANERJEE, A. et al. Clustering with bregman divergences. **Journal of Machine Learning Research**, v. 6, p. 1705–1749, 2005.

_____. Clustering with bregman divergences. **The Journal of Machine Learning Research**, JMLR.org, v. 6, p. 1705–1749, 2005.

BARRON, J. L.; FLEET, D. J.; BEAUCHEMIN, S. S. Performance of optical flow techniques. **International journal of computer vision**, Springer, v. 12, n. 1, p. 43–77, 1994.

BEAULIEU, J.-M.; GOLDBERG, M. Hierarchy in picture segmentation: A stepwise optimization approach. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 11, n. 2, p. 150–163, 1989.

BECK, A.; TEBOULLE, M. Mirror descent and nonlinear projected subgradient methods for convex optimization. **Operations Research Letters**, Elsevier, v. 31, n. 3, p. 167–175, 2003.

BEN-TAL, A.; MARGALIT, T.; NEMIROVSKI, A. The ordered subsets mirror descent optimization method with applications to tomography. **SIAM Journal on Optimization**, SIAM, v. 12, n. 1, p. 79–108, 2001.

BREGMAN, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. **USSR Computational Mathematics and Mathematical Physics**, p. 200–217, 1967.

BROWN, L. G. A survey of image registration techniques. **ACM computing surveys (CSUR)**, ACM, v. 24, n. 4, p. 325–376, 1992.

- BUNTINE, W. Variational extensions to em and multinomial pca. In: **Machine Learning: ECML 2002**. [S.l.]: Springer, 2002. p. 23–34.
- BURGER, M. Bregman distances in inverse problems and partial differential equations. In: **Advances in Mathematical Modeling, Optimization and Optimal Control**. [S.l.]: Springer, 2016. p. 3–33.
- CAHILL, N. D.; NOBLE, J. A.; HAWKES, D. J. Accounting for changing overlap in variational image registration. In: **IEEE. Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on**. [S.l.], 2010. p. 384–387.
- CANNY, J. F. **Finding Edges and Lines in Images**. [S.l.], 1983.
- CAYTON, L. Fast nearest neighbor retrieval for bregman divergences. In: **Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008**. [s.n.], 2008. p. 112–119. Disponível em: <<http://doi.acm.org/10.1145/1390156.1390171>>.
- _____. Efficient bregman range search. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2009. p. 243–251.
- CHEN, C.-S.; YEH, C.-W.; YIN, P.-Y. A novel fourier descriptor based image alignment algorithm for automatic optical inspection. **J. Vis. Comun. Image Represent.**, Academic Press, Inc., Orlando, FL, USA, v. 20, n. 3, p. 178–189, abr. 2009. ISSN 1047-3203. Disponível em: <<http://dx.doi.org/10.1016/j.jvcir.2008.11.003>>.
- CHEN, G.; TEBOULLE, M. Convergence analysis of a proximal-like minimization algorithm using bregman functions. **SIAM Journal on Optimization**, SIAM, v. 3, n. 3, p. 538–543, 1993.
- CHEN, Y.; YE, X. Inverse consistent deformable image registration. In: **The Legacy of Alladi Ramakrishnan in the Mathematical Sciences**. [S.l.]: Springer, 2010. p. 419–440.
- CHIANG, M.-C. et al. Fluid registration of diffusion tensor images using information theory. **Medical Imaging, IEEE Transactions on**, IEEE, v. 27, n. 4, p. 442–456, 2008.
- CHRISTENSEN, G. E.; JOHNSON, H. J. Consistent image registration. **Medical Imaging, IEEE Transactions on**, IEEE, v. 20, n. 7, p. 568–582, 2001.
- CHUMCHOB, N. Vectorial total variation-based regularization for variational image registration. **Image Processing, IEEE Transactions on**, IEEE, v. 22, n. 11, p. 4551–4559, 2013.
- CLARKSON, K. L. Nearest-neighbor searching and metric space dimensions. **Nearest-neighbor methods for learning and vision: theory and practice**, p. 15–59, 2006.
- COSTA, C. L. et al. Mr defecography registration for computer-aided diagnosis of pelvic floor disorder. In: **IEEE. Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on**. [S.l.], 2014. p. 1039–1044.

- COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. New York, NY, USA: Wiley-Interscience, 1991. ISBN 0-471-06259-6.
- COVIELLO, E. et al. That was fast! speeding up nn search of high dimensional distributions. In: **ICML (3)**. [S.l.: s.n.], 2013. p. 468–476.
- CSURKA, G. et al. Visual categorization with bags of keypoints. In: **In Workshop on Statistical Learning in Computer Vision, ECCV**. [S.l.: s.n.], 2004. p. 1–22.
- DASGUPTA, S.; TELGARSKY, M. J. Agglomerative bregman clustering. In: **Proceedings of the 29th International Conference on Machine Learning (ICML-12)**. [S.l.: s.n.], 2012. p. 1527–1534.
- DATTA, R. et al. Image retrieval: Ideas, influences, and trends of the new age. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 40, n. 2, p. 5:1–5:60, maio 2008. ISSN 0360-0300. Disponível em: <<http://doi.acm.org/10.1145/1348246.1348248>>.
- DAVATZIKOS, C. Spatial transformation and registration of brain images using elastically deformable models. **Computer Vision and Image Understanding**, Elsevier, v. 66, n. 2, p. 207–222, 1997.
- DONG, K.; GUO, L.; FU, Q. An adult image detection algorithm based on bag-of-visual-words and text information. In: **10th International Conference on Natural Computation, ICNC 2014, Xiamen, China, August 19-21, 2014**. [s.n.], 2014. p. 556–560. Disponível em: <<http://dx.doi.org/10.1109/ICNC.2014.6975895>>.
- FERREIRA, D. P.; RIBEIRO, E.; BARCELOS, C. A. Variational non rigid registration with bregman divergences. In: **ACM. Proceedings of the Symposium on Applied Computing**. [S.l.], 2017. p. 158–163.
- FERREIRA, D. P.; ROCHA, B. M.; BARCELOS, C. A. Nearest neighbor search on total bregman balls tree. In: **ACM. Proceedings of the Symposium on Applied Computing**. [S.l.], 2017. p. 120–124.
- FERREIRA, D. P. L.; BACKES, A. R.; BARCELOS, C. A. Z. Bregman divergence applied to hierarchical segmentation problems. In: **SPRINGER. Iberoamerican Congress on Pattern Recognition**. [S.l.], 2015. p. 493–500.
- FISCHER, B.; MODERSITZKI, J. A unified approach to fast image registration and a new curvature based registration technique. **Linear Algebra and its applications**, Elsevier, v. 380, p. 107–124, 2004.
- FISCHER, B.; WEBER, H. Express saccades and visual attention. **Behavioral and Brain Sciences**, v. 16, p. 553–567, 9 1993. ISSN 1469-1825. Disponível em: <http://journals.cambridge.org/article_S0140525X00031575>.
- FITZPATRICK, J. M.; HILL, D. L.; JR, C. R. M. Image registration. **Handbook of medical imaging**, v. 2, p. 447–513, 2000.
- FLUSSER, J. Moment invariants in image analysis. In: **proceedings of world academy of science, engineering and technology**. [S.l.: s.n.], 2006. v. 11, n. 2, p. 196–201.

- GAN, R.; CHUNG, A. C.; LIAO, S. Maximum distance-gradient for robust image registration. **Medical image analysis**, Elsevier, v. 12, n. 4, p. 452–468, 2008.
- GAO, Y. et al. Database saliency for fast image retrieval. **IEEE Transactions on Multimedia**, v. 17, n. 3, p. 359–369, 2015. Disponível em: <<http://dx.doi.org/10.1109/TMM.2015.2389616>>.
- GLASBEY, C. A.; MARDIA, K. V. A review of image-warping methods. **Journal of applied statistics**, Taylor & Francis, v. 25, n. 2, p. 155–171, 1998.
- GOLDBERGER, J. et al. Neighbourhood components analysis. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2004. p. 513–520.
- GOLDSTEIN, T.; BRESSON, X.; OSHER, S. Geometric applications of the split bregman method: segmentation and surface reconstruction. **Journal of Scientific Computing**, Springer, v. 45, n. 1-3, p. 272–293, 2010.
- GOSHTASBY, A. A. **Image registration: Principles, tools and methods**. [S.l.]: Springer Science & Business Media, 2012.
- GUETTER, C. et al. Learning based non-rigid multi-modal image registration using kullback-leibler divergence. In: **Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005**. [S.l.]: Springer, 2005. p. 255–262.
- GUO, X.; HONG, J.; YANG, N. Ambiguity set and learning via bregman and wasserstein. **arXiv preprint arXiv:1705.08056**, 2017.
- HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On clustering validation techniques. **Journal of Intelligent Information Systems**, Springer, v. 17, n. 2, p. 107–145, 2001.
- HE, Y.; HAMZA, A. B.; KRIM, H. A generalized divergence measure for robust image registration. **Signal Processing, IEEE Transactions on**, IEEE, v. 51, n. 5, p. 1211–1220, 2003.
- HERMOSILLO, G.; CHEFD'HOTEL, C.; FAUGERAS, O. Variational methods for multimodal image matching. **International Journal of Computer Vision**, Springer, v. 50, n. 3, p. 329–343, 2002.
- JAIN, A. K.; MARTIN, H. Law, data clustering: a user's dilemma. In: **Proceedings of the First international conference on Pattern Recognition and Machine Intelligence**. [S.l.: s.n.], 2005.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: A review. **ACM Comput. Surv.**, ACM, New York, NY, USA, v. 31, n. 3, p. 264–323, set. 1999. Disponível em: <<http://doi.acm.org/10.1145/331499.331504>>.
- KAUFMAN, L.; ROUSSEEUW, P. J. Divisive analysis (program diana). **Finding Groups in Data: An Introduction to Cluster Analysis**, Wiley Online Library, p. 253–279, 1990.
- KIWIEL, K. C. Proximal minimization methods with generalized bregman functions. **SIAM journal on control and optimization**, SIAM, v. 35, n. 4, p. 1142–1168, 1997.

- KLAVA B., H. N. S. T. Segmentação interativa de imagens via transformação watershed. In: SIBGRAPI, W. of (Ed.). **23rd Conference on Graphics, Patterns and Images - SIBGRAPI**. [S.l.: s.n.], 2010. WTD, p. 408 – 413.
- KOEPFLER, G.; LOPEZ, C.; MOREL, J.-M. A multiscale algorithm for image segmentation by variational method. **SIAM journal on numerical analysis**, SIAM, v. 31, n. 1, p. 282–299, 1994.
- KRISTO; CHUA, C. Image representation for object recognition: Utilizing overlapping windows in spatial pyramid matching. In: **IEEE International Conference on Image Processing, ICIP 2013, Melbourne, Australia, September 15-18, 2013**. [s.n.], 2013. p. 3354–3357. Disponível em: <<http://dx.doi.org/10.1109/ICIP.2013.6738691>>.
- KUNZ, T.; BLACK, J. P. Using automatic process clustering for design recovery and distributed debugging. **Software Engineering, IEEE Transactions on**, IEEE, v. 21, n. 6, p. 515–527, 1995.
- LI, B. et al. A new divergence measure based on arimoto entropy for medical image registration. In: **IEEE. Pattern Recognition (ICPR), 2014 22nd International Conference on**. [S.l.], 2014. p. 3197–3202.
- LIU, C.; BELKIN, M. Clustering with bregman divergences: an asymptotic analysis. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2016. p. 2351–2359.
- LIU, J.; VEMURI, B. C.; MARROQUIN, J. L. Local frequency representations for robust multimodal image registration. **Medical Imaging, IEEE Transactions on**, IEEE, v. 21, n. 5, p. 462–469, 2002.
- LIU, J. et al. One step beyond bags of features: Visual categorization using components. In: **18th IEEE International Conference on Image Processing, ICIP 2011, Brussels, Belgium, September 11-14, 2011**. [s.n.], 2011. p. 2417–2420. Disponível em: <<http://dx.doi.org/10.1109/ICIP.2011.6116130>>.
- LIU, M. **Total Bregman Divergence, a Robust Divergence Measure, and Its Applications**. Tese (Doutorado), Gainesville, FL, USA, 2011.
- LIU, M. et al. Shape retrieval using hierarchical total bregman soft clustering. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 34, n. 12, p. 2407–2419, 2012. Disponível em: <<http://dblp.uni-trier.de/db/journals/pami/pami34.html#LiuVAN12>>.
- LIU, W.; RIBEIRO, E. Estimating nonrigid shape deformation using moments. In: **IEEE. Pattern Recognition (ICPR), 2010 20th International Conference on**. [S.l.], 2010. p. 185–188.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, 2004. Disponível em: <<http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>>.
- MAES, F. et al. Image registration using mutual information. In: **Handbook of Biomedical Imaging**. [S.l.]: Springer, 2015. p. 295–308.

MAESSCHALCK, R. D.; JOUAN-RIMBAUD, D.; MASSART, D. L. The mahalanobis distance. **Chemometrics and intelligent laboratory systems**, Elsevier, v. 50, n. 1, p. 1–18, 2000.

MARTIN, S.; DURRANI, T. S. A new divergence measure for medical image registration. **Image Processing, IEEE Transactions on**, IEEE, v. 16, n. 4, p. 957–966, 2007.

MONTGOMERY, D. C.; PECK, E. A.; Vining, G. G. **Introduction to linear regression analysis**. [S.l.]: John Wiley & Sons, 2012. v. 821.

MUMFORD, D.; SHAH, J. Optimal approximations by piecewise smooth functions and associated variational problems. **Communications on pure and applied mathematics**, Wiley Online Library, v. 42, n. 5, p. 577–685, 1989.

MYRONENKO, A.; SONG, X. Intensity-based image registration by minimizing residual complexity. **Medical Imaging, IEEE Transactions on**, IEEE, v. 29, n. 11, p. 1882–1891, 2010.

NAKAMOTO, S.; TORIU, T. Combination way of local properties, classifiers and saliency in bag-of- keypoints approach for generic object recognition. **International Journal of Computer Science and Network Security**, v. 11, 2011.

NG, R. T.; HAN, J. E cient and e ective clustering methods for spatial data mining. In: **Proc. of**. [S.l.: s.n.], 1994. p. 144–155.

NIELSEN, F.; NOCK, R. Total jensen divergences: definition, properties and clustering. In: **IEEE. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.], 2015. p. 2016–2020.

NIELSEN, F.; PIRO, P.; BARLAUD, M. Tailored bregman ball trees for effective nearest neighbors. In: **Proceedings of the 25th European Workshop on Computational Geometry (EuroCG)**. [S.l.: s.n.], 2009. p. 29–32.

NOCK, R.; NIELSEN, F. Bregman divergences and surrogates for learning. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 31, n. 11, p. 2048–2059, 2009. ISSN 0162-8828.

NOCK, R.; NIELSEN, F.; AMARI, S. I. On conformal divergences and their population minimizers. **IEEE Transactions on Information Theory**, v. 62, n. 1, p. 527–538, Jan 2016. ISSN 0018-9448.

OU, Y. et al. Dramms: Deformable registration via attribute matching and mutual-saliency weighting. **Medical image analysis**, Elsevier, v. 15, n. 4, p. 622–639, 2011.

PARK, G.; BAEK, Y.; LEE, H.-K. Re-ranking algorithm using post-retrieval clustering for content-based image retrieval. **Information processing & management**, Elsevier, v. 41, n. 2, p. 177–194, 2005.

PLUIM, J. P. et al. Mutual-information-based registration of medical images: a survey. **Medical Imaging, IEEE Transactions on**, IEEE, v. 22, n. 8, p. 986–1004, 2003.

- PLUIM, J. P.; MAINTZ, J. A.; VIERGEVER, M. A. Interpolation artefacts in mutual information-based image registration. **Computer vision and image understanding**, Elsevier, v. 77, n. 2, p. 211–232, 2000.
- _____. f-information measures in medical image registration. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Medical Imaging 2001**. [S.l.], 2001. p. 579–587.
- POWERS, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. Bioinfo Publications, 2011.
- REED, T. R.; DUBUF, J. H. A review of recent texture segmentation and feature extraction techniques. **CVGIP: Image understanding**, Elsevier, v. 57, n. 3, p. 359–372, 1993.
- ROCHA, B.; FERREIRA, D.; BARCELOS, C. Z. Image retrieval via generalized i-divergence in the bag-of-visual-words framework. In: IEEE. **Electronics, Circuits and Systems (ICECS), 2014 21st IEEE International Conference on**. [S.l.], 2014. p. 734–737.
- ROCHA, B. M. O uso da divergência de kullback-leibler e da divergência generalizada como medida de similaridade em sistemas cbir. Universidade Federal de Uberlândia, 2015.
- ROCKAFELLAR, R. T. **Convex Analysis (Princeton Landmarks in Mathematics and Physics)**. [S.l.]: Princeton University Press, 1996. Paperback.
- ROMERO, A.; LACASSAGNE, L.; GOUIFFES, M. Total bregman divergence for multiple object tracking. In: **IEEE, International Conference on Image Processing (ICIP)**. [S.l.: s.n.], 2013.
- ROOS, J. E. et al. Experience of 4 years with open mr defecography: Pictorial review of anorectal anatomy and disease 1. **Radiographics**, Radiological Society of North America, v. 22, n. 4, p. 817–832, 2002.
- SAI-KIT, Y. et al. Enforcing stochastic inverse consistency in non-rigid image registration and matching. In: IEEE. **Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on**. [S.l.], 2008. p. 1–8.
- SCHWANDER, O.; NIELSEN, F. Reranking with contextual dissimilarity measures from representational bregman k-means. In: **VISAPP (1)**. [S.l.: s.n.], 2010. p. 118–123.
- SHEN, D.; DAVATZIKOS, C. Hammer: hierarchical attribute matching mechanism for elastic registration. **Medical Imaging, IEEE Transactions on**, IEEE, v. 21, n. 11, p. 1421–1439, 2002.
- SOARES, R. de C.; SILVA, I. R. da; GULIATO, D. Spatial locality weighting of features using saliency map with a bag-of-visual-words approach. In: **IEEE 24th International Conference on Tools with Artificial Intelligence, ICTAI 2012, Athens, Greece, November 7-9, 2012**. [s.n.], 2012. p. 1070–1075. Disponível em: <<http://dx.doi.org/10.1109/ICTAI.2012.151>>.

- SOTIRAS, A.; DAVATZIKOS, C.; PARAGIOS, N. Deformable medical image registration: A survey. **Medical Imaging, IEEE Transactions on**, IEEE, v. 32, n. 7, p. 1153–1190, 2013.
- SPERANDIO, D.; MENDES, J. T.; SILVA, L. H. M. e. **Cálculo numérico: características matemáticas e computacionais dos métodos numéricos**. [S.l.]: Prentice Hall, 2003.
- TAN, P.-N. et al. **Introduction to data mining**. [S.l.]: Pearson Education India, 2006.
- TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining, (First Edition)**. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- THIRION, J.-P. Image matching as a diffusion process: an analogy with maxwell's demons. **Medical image analysis**, Elsevier, v. 2, n. 3, p. 243–260, 1998.
- TUCERYAN, M.; JAIN, A. K. et al. Texture analysis. **Handbook of pattern recognition and computer vision**, v. 2, p. 207–248, 1993.
- VENDRAMIN, L.; CAMPELLO, R. J.; HRUSCHKA, E. R. Relative clustering validity criteria: A comparative overview. **Statistical Analysis and Data Mining: The ASA Data Science Journal**, Wiley Online Library, v. 3, n. 4, p. 209–235, 2010.
- WEINBERGER, K. Q.; SAUL, L. K. Distance metric learning for large margin nearest neighbor classification. **The Journal of Machine Learning Research**, JMLR.org, v. 10, p. 207–244, 2009.
- XING, E. P. et al. Distance metric learning with application to clustering with side-information. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2002. p. 505–512.
- XU, B. et al. A bregman divergence optimization framework for ranking on data manifold and its new extensions. In: **AAAI**. [S.l.: s.n.], 2012.
- ZEZULA, P. et al. **Similarity search: the metric space approach**. [S.l.]: Springer Science & Business Media, 2006. v. 32.
- ZHANG, B.; HSU, M.; DAYAL, U. K-harmonic means-a spatial clustering algorithm with boosting. In: **Temporal, Spatial, and Spatio-Temporal Data Mining**. [S.l.]: Springer, 2001. p. 31–45.
- ZHANG, J. et al. Local features and kernels for classification of texture and object categories: A comprehensive study. **International journal of computer vision**, Springer, v. 73, n. 2, p. 213–238, 2007.
- ZHANG, Z.; JIANG, Y.; TSUI, H. Consistent multi-modal non-rigid registration based on a variational approach. **Pattern Recognition Letters**, Elsevier, v. 27, n. 7, p. 715–725, 2006.
- ZHANG, Z. et al. Similarity search on bregman divergence: Towards non-metric indexing. **Proceedings of the VLDB Endowment**, VLDB Endowment, v. 2, n. 1, p. 13–24, 2009.