O uso da Divergência de Kullback-Leibler e da Divergência Generalizada como medida de similaridade em sistemas CBIR

Bruno Moraes Rocha

Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2015

Bruno Moraes Rocha

O uso da Divergência de Kullback-Leibler e da Divergência Generalizada como medida de similaridade em sistemas CBIR

Dissertação de mestrado 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 Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Dra. Celia Aparecida Zorzo Barcelos

Uberlândia
2015

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

R672u 2016

Rocha, Bruno Moraes, 1990-
O uso da Divergência de Kullback-Leibler e da Divergência Generalizada como medida de similaridade em sistemas CBIR / Bruno Moraes Rocha. - 2016.
146 f. : il.

Orientadora: Celia Aparecida Zorzo Barcelos.
Dissertação (mestrado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Ciência da Computação.
Inclui bibliografia.

1. Computação - Teses. 2. Recuperação da informação - Teses. 3. Processamento de imagens - Teses. I. Barcelos, Celia Aparecida Zorzo. II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Ciência da Computação. III. Título.

CDU: 681.3

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 dissertação intitulada "**O uso da Divergência de Kullback-Leibler e da Divergência Generalizada como medida de similaridade em sistemas CBIR**" por **Bruno Moraes Rocha** como parte dos requisitos exigidos para a obtenção do título de **Mestre em Ciência da Computação**.

Uberlândia, 21 de Setembro de 2015

Orientador: _____
Profª. Dra. Celia Aparecida Zorzo Barcelos
Universidade Federal de Uberlândia

Banca Examinadora:

_____
Profª. Dra. Denise Guliatto
Universidade Federal de Uberlândia

_____
Prof. Dr. Marcos Aurélio Batista
Universidade Federal de Goiás

A Deus, que nos criou com seu imenso poder. Seu fôlego de vida em mim foi como sustento. Sua mão me guardando me forneceu a coragem e determinação necessária para nunca temer os desafios encontrados.

Agradecimentos

Quero expressar a minha gratidão e reconhecimento a Deus, pessoas e entidades que colaboraram com a realização deste trabalho.

Primeiramente agradeço a Deus, por me ajudar a superar todos os obstáculos encontrados durante a minha trajetória e estar sempre realizando grandes milagres em minha vida. E, por este, não é possível descrever os meus sentimentos.

Agradeço aos meus pais, Sandra Aparecida de Moraes Lima e Diomar Bento da Rocha, pelos cuidados, sentimento e o amor, por deles recebido. Também ao meu tio Álvaro de Moraes Lima, por me ensinar a não desistir, independentemente da dificuldade encontrada.

Ao meu irmão Hudson Moraes Rocha, que sempre esteve disposto a me ajudar e que, também foi motivo de inspiração para almejar os meus objetivos. Em especial, à minha cunhada Walquíria Dutra de Oliveira, que foi paciente e amiga, que me ajudou com muita dedicação na conclusão da minha dissertação.

À minha noiva, Emília Alves Nogueira, pela companhia, carinho, motivação, amor e força.

Aos meus amigos Reny, Marcos Aurélio, Edmundo, Dirles e Juan, por sempre se mostraram presentes em minha vida no momento que havia uma grande distância entre nós.

A minha orientadora, professora Dra. Celia Aparecida Zorzo Barcelos, pela orientação, profissionalismo e a experiência científica transmitida.

Aos meus colegas Franciny, Walter, Reslley, Joyce, Juan, Myllene, Maria Cristina, Acrísio e Rafael, turma na qual ingressei na pós-graduação e sempre estudamos juntos para conseguir cumprir os créditos necessários para o programa da pós-graduação.

Também ao Ernani e à Daniela, que compartilharam suas experiências e considerações, para auxiliar na conclusão do trabalho de mestrado.

À infra-estrutura da Universidade Federal de Uberlândia.

Agradeço à CAPES, pelo auxílio que viabilizou o desenvolvimento desta pesquisa e permitiu a minha dedicação exclusiva.

"Deus nos abençoou com muitas coisas, principalmente o livre arbítrio,
caso não soubermos como usá-lo, o destino será incerto."
(Anônimo)

Resumo

A recuperação de imagem baseada em conteúdo é importante para diversos fins, como diagnósticos de doenças a partir de tomografias computadorizadas, por exemplo. A relevância social e econômica de sistemas de recuperação de imagens criou a necessidade do seu aprimoramento. Dentro deste contexto, os sistemas de recuperação de imagens baseadas em conteúdo são compostos de duas etapas: extração de característica e medida de similaridade. A etapa de similaridade ainda é um desafio, devido à grande variedade de funções de medida de similaridade, que podem ser combinadas com as diferentes técnicas presentes no processo de recuperação e retornar resultados que nem sempre são os mais satisfatórios. As funções geralmente mais usadas para medir a similaridade são as Euclidiana e Cosseno, mas alguns pesquisadores têm notado algumas limitações nestas funções de proximidade convencionais, na etapa de busca por similaridade. Por esse motivo, as divergências de Bregman (Kullback Leibler e Generalizada) têm atraído a atenção dos pesquisadores, devido à sua flexibilidade em análise de similaridade. Desta forma, o objetivo desta pesquisa foi realizar um estudo comparativo sobre a utilização das divergências de Bregman em relação às funções Euclidiana e Cosseno, na etapa de similaridade da recuperação de imagens baseadas em conteúdo, averiguando as vantagens e desvantagens de cada função. Para isso, criou-se um sistema de recuperação de imagens baseado em conteúdo em duas etapas: *off-line* e *on-line*, utilizando as abordagens BSM, FISM, BoVW e BoVW-SPM. Com esse sistema, foram realizados três grupos de experimentos utilizando os bancos de dados: Caltech101, Oxford e UK-bench. O desempenho do sistema de recuperação de imagem baseada em conteúdo utilizando as diferentes funções de similaridade foram testadas por meio das medidas de avaliação: *Mean Average Precision*, *normalized Discounted Cumulative Gain*, precisão em k, e precisão x revocação. Por fim, o presente estudo aponta que o uso das divergências de Bregman (Kullback Leibler e Generalizada) obtiveram melhores resultados do que as medidas Euclidiana e Cosseno, com ganhos relevantes para recuperação de imagem baseada em conteúdo.

Palavras-chave: I-Divergence Generalizada. Kullback Leibler. Similaridade. Divergência de Bregman. Recuperação.

Abstract

The content-based image retrieval is important for various purposes like disease diagnoses from computerized tomography, for example. The relevance, social and economic of image retrieval systems has created the necessity of its improvement. Within this context, the content-based image retrieval systems are composed of two stages, the feature extraction and similarity measurement. The stage of similarity is still a challenge due to the wide variety of similarity measurement functions, which can be combined with the different techniques present in the recovery process and return results that aren't always the most satisfactory. The most common functions used to measure the similarity are the Euclidean and Cosine, but some researchers have noted some limitations in these functions conventional proximity, in the step of search by similarity. For that reason, the Bregman divergences (Kullback Leibler and I-Generalized) have attracted the attention of researchers, due to its flexibility in the similarity analysis. Thus, the aim of this research was to conduct a comparative study over the use of Bregman divergences in relation the Euclidean and Cosine functions, in the step similarity of content-based image retrieval, checking the advantages and disadvantages of each function. For this, it was created a content-based image retrieval system in two stages: offline and online, using approaches BSM, FISM, BoVW and BoVW-SPM. With this system was created three groups of experiments using databases: Caltech101, Oxford and UK-bench. The performance of content-based image retrieval system using the different functions of similarity was tested through of evaluation measures: Mean Average Precision, normalized Discounted Cumulative Gain, precision at k, precision x recall. Finally, this study shows that the use of Bregman divergences (Kullback Leibler and Generalized) obtains better results than the Euclidean and Cosine measures with significant gains for content-based image retrieval.

Keywords: Generalized I-Divergence. Kullback Leibler. Similarity. Bregman divergence. Retrieval.

Lista de ilustrações

Lista de tabelas

Lista de siglas

DT-CWT *Dual-Tree Complex Wavelet Transform* 83

DWT *Discrete Wavelet Transform* 83

DMR *divergence view of MR* 77

FISM *Fuzzy Image Descriptor Based on Saliency MAP* 16

FE *feature extraction* 83

GID *Generalized I-Divergence* 16

IR *Information Retrieval* 49

idd *independent and identically distributed* 83

KL Kullback Leibler 30

KLS Kullback Leibler Simétrica 79

k-NNq *k-Nearest Neighbors Query* 46

k-NN *k-Nearest Neighbors* 76

LS *Lifting Schemes* 79

MAP *Mean Average Precision* 32

MS mapa de saliência 60

MR *manifold ranking* 75

MLE *maximum-likelihood estimators* 85

nDCG *normalized Discounted Cumulative Gain* 32

NED *Normalized Euclidean Distance* 81

PCA *Principal Component Analysis* 78

PCA-SIFT *Principal Component Analysis*-SIFT 55

PDF *Probability Density Function* 83

RGB *Red Green Blue* 38

RQ *Range Query* 46

SIFT *Scale Invariant Feature Transform* 15

SPM *Spatial Pyramids Matching* 37

S-SIFT *Sparse Scale-Invariant Feature Transform* 58

TI Teoria da Informação 89

TCN Teoria dos Conjuntos Nebulosos 65

WT *Wavelet Transform* 75

Sumário

Capítulo 1

Introdução

O avanço da tecnologia digital e a diminuição do custo de armazenamento de dados contribuíram para um aumento do número de imagens: na *Internet*, em banco de dados públicos e em sistemas biométricos. Nesse sentido, vários sistemas de recuperação de imagens têm sido desenvolvidos na tentativa de otimizar a consulta do usuário a esses bancos de dados. Os sistemas computacionais de recuperação de imagens são baseados em duas técnicas de busca, uma em texto e outra em conteúdo.

Nas técnicas de busca por texto, o processo de recuperação de imagens consiste em comparar os termos de uma consulta textual, definida por um usuário, com as anotações associadas, às imagens, por exemplo, representadas por palavras-chave e, a partir da comparação, retornar ao usuário um conjunto de imagens. Entretanto, essa técnica apresenta duas desvantagens: a primeira é a necessidade de um trabalho manual para realizar as anotações e, a segunda, refere-se às incertezas das palavras usadas para a recuperação de imagens (MÜLLER et al., 2004).

As técnicas de recuperação de imagens baseadas em conteúdo (*Content Based Image Retrieval* – CBIR) (IQBAL et al., 2014), têm sido propostas na tentativa de superar as desvantagens de sistemas de recuperação de imagens baseados em texto (SNOEK; SMEULDERS, 2010). Nessa abordagem, são consideradas as informações visuais da imagem para a busca e recuperação em um banco de dados e, não apenas uma simples descrição textual das mesmas (BALAN et al., 2004).

Os sistemas CBIR têm ganhado relevância, principalmente, pela subjetividade em se caracterizar uma imagem pelo seu conteúdo, já que diferentes usuários podem estar interessados em diferentes aspectos de uma mesma imagem (BALAN et al., 2004). Em (MARQUES et al., 2002), por exemplo, os autores implementaram um sistema CBIR para análise de imagens de mamograma, com intuito de averiguar a presença de microcalcificações nas imagens de mamografias para possíveis diagnóstico de casos iniciais de câncer de mama. O trabalho de (TORRES; FALCÃO, 2006), propõe um sistema CBIR na área de biodiversidade, para auxiliar a identificação de espécies de animais por meio de suas formas.

Entretanto, mesmo com todos os esforços nas pesquisas de recuperação de imagem baseada em conteúdo, os algoritmos atuais de CBIR ainda são limitados (SILVA, 2014). Além de outras dificuldades, o gargalo principal é a descontinuidade existentes entre os seus conteúdos semânticos associados e as características de baixo nível possíveis a serem extraídas (DATTA et al., 2008). A descontinuidade semântica é um problema originado do fato que medidas de similaridade e os extratores de caraterísticas das imagens, tais como histogramas de níveis de cinza, descritores de forma e cor, não possuem ligação direta com as semânticas da subjetividade humana (DESERNO; WELTER; HORSCH, 2012).

Visando minimizar o problema semântico, diversos trabalhos têm abordado CBIR com diferentes medidas de similaridade (SCHOLAR, 2013; ABOOD; MUHSIN; TAWFIQ, 2013; KEKRE; SONAWANE, 2012). A proposta deste trabalho insere-se neste contexto, propondo o uso das divergências de Bregman Kullback Leibler (KL) e GID como medida de similaridade em CBIR, na etapa de recuperação de imagem. A relevância desta proposta está ligada à possibilidade de estabelecer a similaridade de forma mais eficaz, visto que estas divergências apresentam propriedades que permitem minimizar os problemas descritos anteriormente.

1.1 Motivação

Os mecanismos de recuperação de imagens baseados em conteúdo têm o seguinte funcionamento: um usuário define uma imagem de consulta (*query*), compara esta imagem com as imagens do banco de dados e retorna uma lista ranqueada contendo as imagens mais similares.

Os sistemas CBIR são baseados em duas etapas principais: a primeira consiste na extração de características, enquanto que a segunda, na medida de similaridade. A extração de característica é o processo no qual um conjunto de características é gerado para representar o conteúdo de cada imagem. Existem vários métodos de extração de características e algumas das mais populares são extrações baseadas em cor, textura e forma. Na etapa da medida de similaridade, uma etapa posterior à extração de característica, aplica-se uma função de distância (por exemplo, Euclidiana) entre os vetores de características da imagem de consulta e de cada uma das imagens que estão no banco de dados, com o intuito de obter a recuperação das N imagens mais semelhantes contidas no banco de dados.

Tanto o processo de extração de característica quanto a medida de similaridade representam um desafio para sistemas CBIR. Considerando a etapa de extração de características, o desafio é a utilização de descritores que possibilitem a minimização da diferença entre as concepções semânticas de alto nível, utilizadas pelos humanos para compreender o conteúdo de uma imagem, e as características de baixo nível, usadas na visão computa-

cional, denominada de *gap*-semântico. Uma possível solução seria o desenvolvimento de algoritmos sofisticados para extração de característica.

A variedade das medidas de similaridade encontradas na literatura tal como: Euclidiana, Mahalanobis e Cosseno (ABOOD; MUHSIN; TAWFIQ, 2013; YANG; XIAO, 2008; ZHOU; DAI, 2006; SPERTUS; SAHAMI; BUYUKKOKTEN, 2005; SANTINI; JAIN, 1999); e as diferentes técnicas de recuperação de imagem dificultam a escolha da medida mais adequada na recuperação de imagens em sistemas CBIR. É importante observar que as medidas de similaridade escolhidas devem ser apropriadas com diferentes técnicas presentes no processo de recuperação de imagens. Por exemplo: (LIU et al., 2008) ao utilizar a distância *City Block* como medida de similaridade em um sistema CBIR os autores não obtiveram bons resultados. Entretanto, (KEKRE; SONAWANE, 2012) usavam a distância Minkowski para elaborar uma seleção de medida de similaridade adequada de acordo com os métodos presentes (por exemplo, extração de características) no CBIR. A utilização de diferentes distâncias para um mesmo processo de recuperação pelos pesquisadores, como os citados acima, denotam a dificuldade em se definir a melhor medida de similaridade a ser usada na recuperação de imagens em sistemas CBIR.

Observa-se ainda que as funções de proximidade convencionais, tais como a Euclidiana e a Cosseno, têm apresentado limitações na busca por similaridades (XU et al., 2012; LIU, 2011; SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010). Diante destas limitações e considerando as suas propriedades, as divergências de Bregman têm sido utilizadas em diversas aplicações como medida de similaridade. Por exemplo, (CAYTON, 2008) propõe uma forma eficiente de encontrar os vizinhos mais próximos utilizando a divergência de Bregman (DB), e (BANERJEE et al., 2005) apresenta uma análise paramétrica *hard* e *soft* de algoritmos de agrupamentos baseados nas DB's.

Desta forma, acredita-se que a utilização das DB's, devido à sua flexibilidade em relação às outras medidas (Euclidiana e Cosseno), podem ser mais eficazes para o cálculo da similaridade entre diferentes características que representam as imagens. As DB's utilizadas neste trabalho são a KL e a GID, as quais são definidas utilizando a função logarítmica cujo domínio é $x > 0$; por outro lado, os dados caracterizados podem assumir valores iguais a zero em suas coordenadas. Nesta perspectiva, são apresentados neste trabalho tratamentos adequados que possibilitam a aplicação das divergências KL e GID, minimizando os problemas descritos anteriormente para recuperação de imagens baseados em conteúdo.

1.2 Objetivos

A presente pesquisa teve por objetivo geral criar tratamentos adequados para as divergências de Bregman (KL e GID), quando as representações das imagens contém coordenadas iguais a 0 (zero), realizando um estudo comparativo sobre o uso dos tratamentos

em relação às medidas Euclidiana e Cosseno, na etapa de similaridade da recuperação de imagens baseadas em conteúdo, verificando as vantagens e desvantagens de cada função.

Os **objetivos específicos** foram:

1. Avaliar o desempenho das divergências de Bregman (KL e GID), na etapa de recuperação de imagens, utilizando os métodos de avaliação precisão e revocação, *normalized Discounted Cumulative Gain* (nDCG), *Mean Average Precision* (MAP) e precisão em k;

2. Comparar o desempenho da CBIR, utilizando a divergências (KL e GID) e as medidas Euclidiana e Cosseno para o cálculo de similaridade na etapa de recuperação;

3. Desenvolver tratamentos para as divergências de Bregman (KL e GID) – obedecendo às propriedades da função logarítmica cujo domínio é $x > 0$, de acordo com a caracterização das imagens;

Com base nos objetivos descritos acima, foram levantadas as hipóteses destacadas na Seção 1.3 abaixo.

1.3 Hipótese

O uso das divergências de Bregman (KL e GID) na etapa de similaridade quando usado histogramas é mais eficaz na recuperação de imagens baseada em conteúdo do que as medidas tradicionais (Euclidiana e Cosseno), dependendo dos tratamentos aplicados para assegurar o domínio das funções logarítmicas.

1.4 Contribuições

Podem-se destacar duas principais contribuições deste trabalho. A primeira contribuição é o estudo comparativo entre as divergências de Bregman (KL e GID) e as medidas tradicionais (Euclidiana e Cosseno), discriminando as vantagens e desvantagens da utilização de cada uma dessas funções. Em segundo lugar, destacam-se os tratamentos para a utilização das divergências de Bregman, com o propósito de minimizar os problemas enfrentados na recuperação de imagens.

1.5 Organização do documento

Para exposição do estudo, e visando uma melhor compreensão dos conceitos e técnicas utilizadas, dos experimentos realizados e dos resultados obtidos, subdividimos este texto em 8 capítulos.

- Capítulo 1 consiste das considerações iniciais e do contexto no qual se insere esta pesquisa, da motivação para o seu desenvolvimento e dos objetivos a serem alcançados.
- Capítulo 2 apresenta uma breve exposição dos conceitos básicos ligados ao processo de recuperação de imagens, enfatizando a recuperação de imagens baseadas em conteúdo. Posteriormente, serão detalhadas as fases e processos envolvidos na recuperação de imagens baseadas em conteúdo.
- Capítulo 3 tece sobre algumas abordagens utilizadas para caracterizar um conjunto de imagens.
- Capítulo 4 define a DB e descreve sobre três trabalhos relacionados à utilização da divergência de Bregman no contexto de recuperação.
- Capítulo 5, descreve a proposta realizada neste estudo
- Capítulo 6 detalha a metodologia empregada, bem como suas etapas, técnicas e recursos utilizados para realizar os experimentos.
- Capítulo 7, apresenta os métodos de avaliação utilizados para esses estudo e discorremos sobre os resultados encontrados na realização dos experimentos deste trabalho.
- Capítulo 8, tecemos as considerações finais, uma síntese das principais contribuições do presente trabalho e propostas para pesquisas futuras.

Capítulo 2

Fundamentação Teórica

Este capítulo inicialmente aborda os conceitos fundamentais utilizados nos sistemas de recuperação de imagens e, especificamente, nos sistemas de recuperação de imagens baseada em conteúdo (*Content-Based Image Retrieval* – CBIR). Por fim, é apresentada alguns métodos para avaliação dos sistemas CBIR.

2.1 Sistema de Recuperação de Imagens: Conceitos Gerais

Os primeiros trabalhos publicados sobre a recuperação de imagens foram na década de 70, e suas técnicas eram baseadas em anotações textuais da imagem (FENG; SIU; ZHANG, 2003). Em outras palavras, as imagens eram descritas com textos, os quais eram utilizados para busca em sistemas de gerenciamento de banco de dados. Porém, a descrição textual das imagens não era uma tarefa trivial (FENG; SIU; ZHANG, 2003), pois as anotações textuais da imagem eram feitas de forma manual, tornando-se uma tarefa complicada e cansativa, principalmente em grandes bases de dados. Os textos não conseguem representar semanticamente todo o conteúdo da imagem. Além disso, a recuperação baseada em texto tinha dificuldade em relacionar o texto da consulta (solicitação do usuário) às anotações da imagem.

No inicio de 1990, as novas tecnologias de sensores de imagem digital possibilitaram o aumento do volume de imagens digitais produzidas por indústrias, áreas médicas, e outras aplicações disponíveis para usuários. Assim, o gerenciamento eficiente para a rápida expansão da informação visual tornou-se um problema a ser resolvido. Diante das dificuldades enfrentadas pela recuperação em textos fez-se necessária a busca e aplicação de outras formas de recuperação da informação, além da baseada em texto (FENG; SIU; ZHANG, 2003).

Em 1992, surgiram novas direções para os sistemas de gerenciamento de banco de dados de imagens. Uma das novas maneiras foi a recuperação baseada nas propriedades

inerentes ao conteúdo da imagem, pois se tornava mais eficiente e intuitivo representar e recuperar a imagem. Desde então, pesquisadores da área computacional, de gerenciamento de banco de dados, e de recuperação de informação, têm sido atraídos por este campo (DAS; MANMATHA; RISEMAN, 1999).

O processo geral de recuperação de imagens pode ser definido como uma pesquisa especializada em banco de dados para encontrar imagens relevantes conforme a requisição do usuário (consulta). A consulta do usuário pode ser por: palavras-chave, arquivo de imagem, ou clique em alguma imagem. O critério para estabelecer se a imagem é relevante dependerá da medida de similaridade entre os documentos do banco em relação à consulta do usuário, e da forma de representação das características das imagens. As medidas de similaridade (também chamadas aqui de funções de similaridade ou função de distância) são baseadas na similaridade "distância" entre as representações quantitativas das características de imagem ou a associação de palavras-chave.

Assim, temos que os sistemas de recuperação de imagens, como explicitado anteriormente, podem utilizar duas abordagens diferentes, que são: baseada em anotações textuais ou baseada no conteúdo visual (também chamada de recuperação de imagem baseada em conteúdo). Enfatizamos que o foco deste trabalho não é a recuperação de imagens baseada em anotações textuais (maiores informações sobre este assunto podem ser encontradas em (CHANG; HSU, 1992) e (TAMURA; YOKOYA, 1984)). A seguir, explanamos sobre a recuperação de imagem baseada em conteúdo visual, foco desta pesquisa.

2.2 Recuperação de Imagens Baseada em Conteúdo

A recuperação de imagens baseada em conteúdo utiliza o conteúdo visual de uma imagem, tal como cor, forma, textura e organização espacial, para representar e indexar a imagem. Em um sistema *Content-Based Image Retrieval* (CBIR), o conteúdo visual das imagens do banco de dados são extraídas e descritas por vetores de características multi-dimensionais, ou seja, o vetor de características das imagens em um banco de dados formam um novo banco de características. Para recuperar as imagens, o usuário fornece uma imagem ou parte dela para consulta e o sistema, então, representa esta imagem por um vetor de característica. Posteriormente, são calculadas as medidas de similaridades entre o vetor de característica da consulta e os vetores do banco de características e a recuperação é realizada com o auxílio de um sistema de indexação para relacionar o vetor de característica com a imagem correspondente. A Figura 1, a seguir, apresenta o processo geral de recuperação de imagens por conteúdo, segundo (FENG; SIU; ZHANG, 2003).

Um sistema CBIR pode ser divido em duas etapas principais: caracterização e similaridade. Na primeira etapa, denominada de caracterização ou extração de características, têm-se a representação de uma imagem por meio da indexação desta por um vetor de característica, derivado do seu conteúdo visual. Já, na segunda etapa da recuperação de

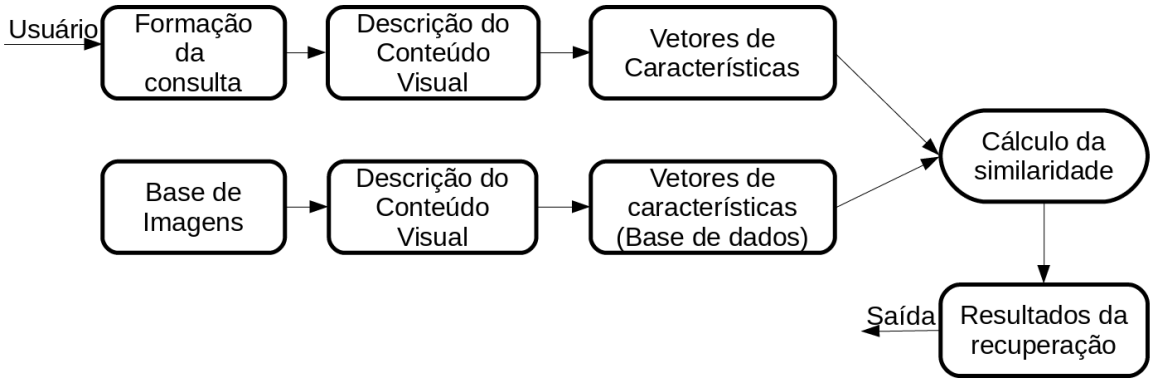

Figura 1 – Fluxo de funcionamento de um sistema CBIR.

Modificada de: (FENG; SIU; ZHANG, 2003).

imagens baseadas em conteúdo, denominada de similaridade, são realizadas outras duas etapas, a medida de similaridade e a consulta por similaridade.

O objetivo do cálculo de similaridade é definir a semelhança das imagens da base de dados em relação à imagem de consulta e, ao final, retornar ao usuário uma lista ranqueada das imagens do banco de dados, na qual as primeiras imagens do ranqueamento são as mais similares com a imagem de consulta (FENG; SIU; ZHANG, 2003). A consulta de similaridade objetiva procurar elementos em um conjunto, segundo algum critério de similaridade, que sejam mais "parecidos" ou mais "distintos" com um outro determinado elemento.

Além disso, os componentes que formam um sistema CBIR adotam, também, o método *Bag-of-Visual-Words* (BoVW) que emprega um conjunto de abordagens para representar uma imagem e pode ser combinado com diferentes técnicas que utilizam a atenção visual, como por exemplo, as abordagens BSM e o FISM, e as pirâmides espaciais *Spatial Pyramids Matching* (SPM) (Seção 3.3).

Nesse contexto, para calcular a similaridade dos vetores de características em um sistema CBIR, geralmente, utilizam-se medidas tradicionais, como por exemplo a Euclidiana e Cosseno. Detalhes dessas distâncias estão abordados no tópico 2.2.2.1. Entretanto (LIU et al., 2012) e (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010), trabalham com as divergências de Bregman que, ao contrário das mencionadas anteriormente (chamaremos aqui de distâncias tradicionais), apresentam uma maior flexibilidade na análise de similaridade. As divergências de Bregman são apresentadas no Capítulo 4.

Assim, na próxima seção, detalharemos as etapas principais desse sistema: a caracterização e a similaridade.

2.2.1 Caracterização ou Extração de características

A extração de características ou caracterização é um processo que compõe um sistema CBIR, no qual se obtém a representação de uma imagem por meio de suas propriedades. Uma imagem é a representação visual de um objeto por meio de alguma técnica como, por exemplo, uma pintura, uma fotografia, desenhos, vídeos, dentre outros. Cada imagem tem muita subjetividade relacionada a ela, o que torna difícil sua caracterização.

Uma imagem pode ser considerada como uma função bidimensional $f(a, b)$ onde a e b são coordenadas planas, e a amplitude de f em qualquer par de coordenadas (a, b) é uma única amostra de um espaço de cores que, tipicamente, são compostas com tons de cinza, variando entre o preto como a intensidade mais baixa e o branco com a intensidade maior, normalmente denominada de intensidade ou nível de cinza da imagem no referido par de coordenadas. Quando (a, b) e a amplitude f fazem parte de um conjunto de valores finitos, discretos, a imagem é denominada de imagem digital (GONZALEZ; WOODS, 2006). As imagens coloridas possuem mais de uma banda de frequência e devem ser representadas por mais de uma função $f(a, b)$, um exemplo deste modelo de cores é o *Red Green Blue* (RGB) que apresenta uma função de intensidade para cada cor primária (vermelho, verde e azul).

Assim, podemos definir uma imagem da seguinte forma: seja uma imagem $I : \Omega \subset \mathbb{R}^S \rightarrow [a, b] \subset \mathbb{R}$ tal que $x \mapsto I(x)$ onde Ω é o suporte da imagem ou $dom(I) = \Omega$ e $S = 2$. Uma imagem em $\mathbb{R}^S$ tem $n \times m$ elementos, podendo também ser representada por um vetor de característica da seguinte maneira, $I \cong (x_1, x_2, ..., x_p)$, onde $p \leq n \times m$.

O processo de caracterização viabiliza extrair automaticamente[1] vetores de características das imagens, por meio dos descritores. A utilização de vetores de características na recuperação por conteúdo, ao invés das imagens propriamente ditas, trazem vantagens como: redução de dimensionalidade, a diminuição do custo computacional e a representação principal do conteúdo da imagem de acordo com o descritor escolhido.

O processo de extração de características deveria ter a capacidade de extrair as características relevantes de uma dada imagem de maneira similar a de um observador humano, entretanto, essa operacionalidade ainda não foi alcançada, devido à limitação do conhecimento científico referente à visão, cognição e a emoção humana (BUGATTI, 2012).

Por isso, as características, de baixo nível, que melhor satisfazem ao critério de separabilidade de imagens, tanto por humanos e pelas máquinas são: cor, forma e textura. A seguir, apresentamos os principais descritores utilizados para representar as características visuais, de baixo nível, de imagens.

[1] Significa que não há nenhuma intervenção humana durante o processo de extração de características.

2.2.1.1 Cor

A propriedade de cor é a característica visual mais utilizada em sistemas CBIR (*content-based image retrieval*), devido ao seu baixo custo computacional. Os extratores de características de cor baseiam-se, principalmente, em histogramas, que foram introduzidos por (SWAIN; BALLARD, 1991), mas que não contêm informações sobre a distribuição espacial de cor, pois se baseiam na frequência das cores. O algoritmo de extração do histograma de cor pode ser dividido nos seguintes passos: (1) particionamento do espaço de cores em células; e (2) armazenam toda a contagem da cor no compartimento do histograma correspondente.

Os valores dos histogramas de cores podem ser normalizados e apresentam algumas vantagens que são: a eficiência de sua computação e a invariância das propriedades de rotação, escala e translação nas imagens. Se a imagem for em tons de cinza, podem-se obter os histogramas de níveis de cinza considerando a iluminação e a saturação (GRUNDLAND; DODGSON, 2007). A Figura 2, a seguir, apresenta um exemplo de dois histogramas de cores com suas respectivas imagens quantizadas em 256 níveis de cinza.

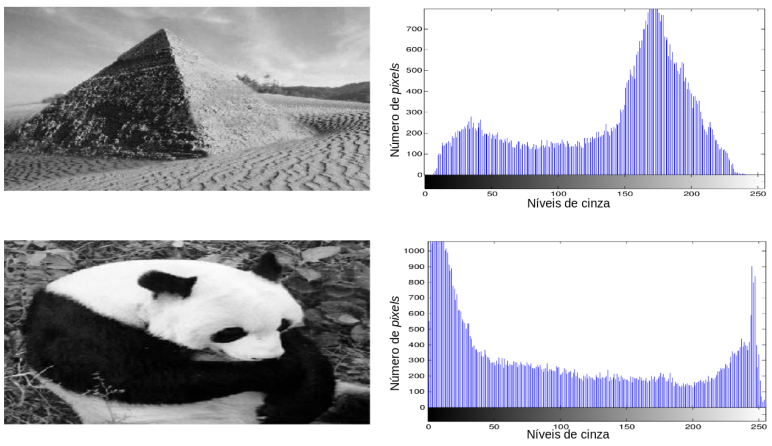

Figura 2 – Do lado esquerdo tem as imagens em 256 níveis de cinza e do lado direito os respectivos histogramas de níveis de cinza da imagem.

Entretanto, a falta de informações sobre a distribuição espacial das cores no histograma, torna-se uma desvantagem (SWAIN; BALLARD, 1991), pois faz com que as imagens muito diferentes tenham representações semelhantes (KIMURA et al., 2011) como representado na Figura 3, a seguir.

Uma outra desvantagem de histograma de cores é sua alta dimensionalidade, também chamada de a "maldição da alta dimensionalidade" (*dimensionality curse*) (J.HERRMANN; P.FRIEDLANDER; YILMAZ, 2012), que é a dimensionalidade (tamanho) do vetor de característica, normalmente de ordem 10^2.

Para solucionar esse problema de informações espaciais das cores no histograma, foram desenvolvidas algumas técnicas que utilizam descritores de cor locais que incluem

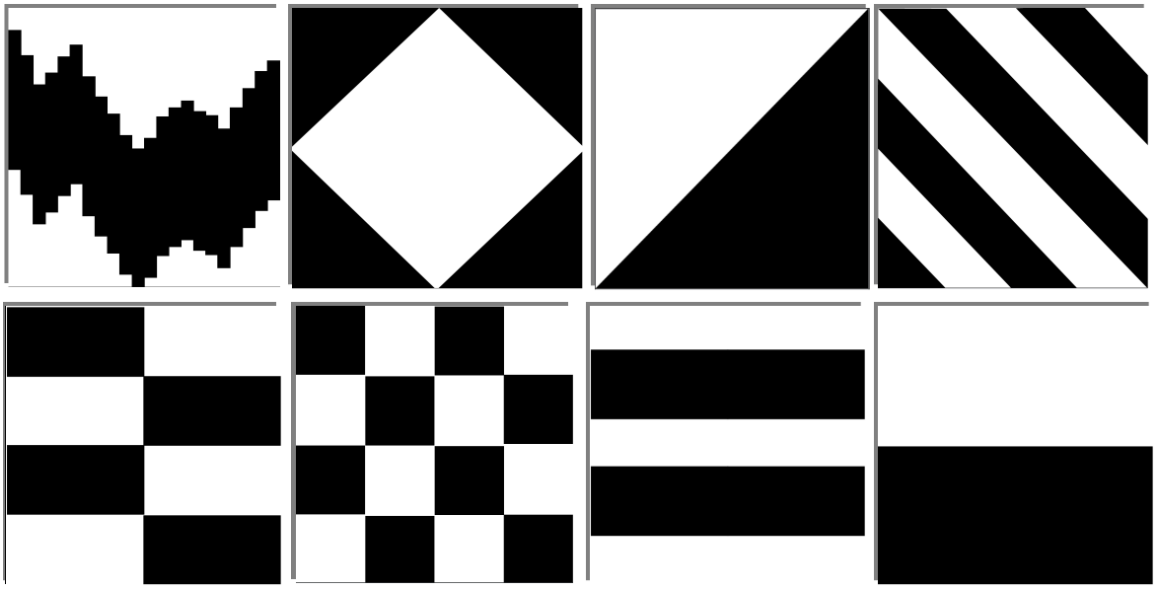

Figura 3 – Imagens visualmente diferentes mas com histogramas equivalentes.

informação espacial sobre o conteúdo visual de imagens. Descritores de cor locais podem ser classificados em dois grupos: abordagens baseadas em partições e abordagens de regiões. As abordagens baseadas em partições incluem informação espacial de características visuais particionando a imagem em blocos de tamanho fixo e em seguida extraindo as características de cada bloco individualmente, o mesmo esquema de partição é aplicado para todas as imagens. Por outro lado, as abordagens de regiões dividem a imagem em regiões, que podem ser de tamanhos distintos para cada imagem.

Além do histograma de cor existem outros descritores de cores como: *Color and Edge Directivity Descriptor* (BAMPIS et al., 2015), *Border Interior Pixel Classification* (IMRAN; HASHIM; KHALID, 2014) *Spatial Configuration of Dominant Color Regions* (JANG; HAN; KIM, 2014), *Color Coherence Vector* (SIDRAM; BHAJANTRI, 2012), *Color Correlogram* (ZHAO; WANG; KHAN, 2011), *Cell Histograms*, dentre outros. A seguir, apresentamos o segundo descritor de extração de características de uma imagem, utilizando o atributo Forma.

2.2.1.2 Forma

O atributo forma é considerado um dos melhores atributos para se representar e identificar um objeto, e mesmo sendo um dos atributos mais difíceis de se caracterizar, especialmente pelo fato de ser necessário segmentar os objetos de interesse contidos na imagem, têm sido utilizado em vários sistemas de recuperação de imagens (LONCARIC, 1998) e (ZHANG; LU, 2004).

Os descritores de formas são baseados em métodos de contorno e em regiões (ZHANG; LU, 2004) que levam em consideração as características extraídas dos contornos ou da

região inteira. Os descritores baseados em regiões expressam a distribuição dos *pixels* como uma região de um objeto, permitindo descrever objetos mais complexos com múltiplas regiões desconexas e/ou objetos conexos que contêm ou não buracos. Já os descritores baseados em contornos apresentam as propriedades da forma pelo seu esboço (contorno), considerando as delimitações (fronteiras) mais externas do objeto, como exemplificado na Figura 4, em sequência.

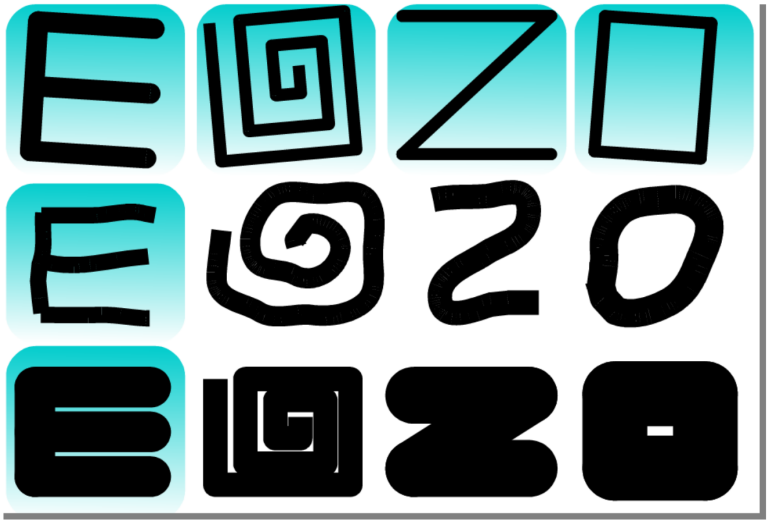

Figura 4 – Exemplos de similaridade de forma baseada em contorno e região.

Na Figura 4, verifica-se que os objetos situados na última linha possuem uma distribuição de *pixels* semelhantes se utilizarmos o método baseado em regiões, enquanto que, se utilizarmos o método baseado em contorno, verificamos que os objetos localizados na última linha seriam diferentes entre si, já que seriam mais semelhantes aos outros objetos no mesmo alinhamento em coluna.

Desta forma, os extratores de características de forma podem ser considerados simples assinaturas do contorno de objetos contidos nas imagens, ou também, como sofisticados descritores baseados em contorno como, por exemplo, saliência de contornos utilizada por (TORRES; FALCÃO, 2007) e *Tensor Scale* utilizado por (ANDALÓ et al., 2010).

Por fim, em sequência, apresentamos a característica de baixo nível, denominada de Textura, e seus respectivos descritores.

2.2.1.3 Textura

A característica textura tem um papel tão importante quanto a de cor e forma na recuperação de imagens baseada em conteúdo. A textura ocorre sobre uma região em vez de um único ponto (*pixel*), manifestando diversos padrões resultantes das propriedades físicas da superfície dos objetos como: aspereza, granularidade, homogeneidade, contraste, rugosidade, direção, resultado de diferenças de reflexão pela absorção ou não da luz na superfície, dentre outros. Essas características complexas da textura a torna interessante

e bastante aplicada às imagens, gerando uma diversidade de métodos para a extração de características.

A textura é uma característica altamente discriminante pelo sistema visual humano, enquanto que, para sistemas automáticos, essa tarefa é mais complicada e necessita de algoritmos mais complexos. Essa característica está presente em quase todos os lugares, de formas distintas e em diferentes ambientes, conforme apresentado na Figura 5.

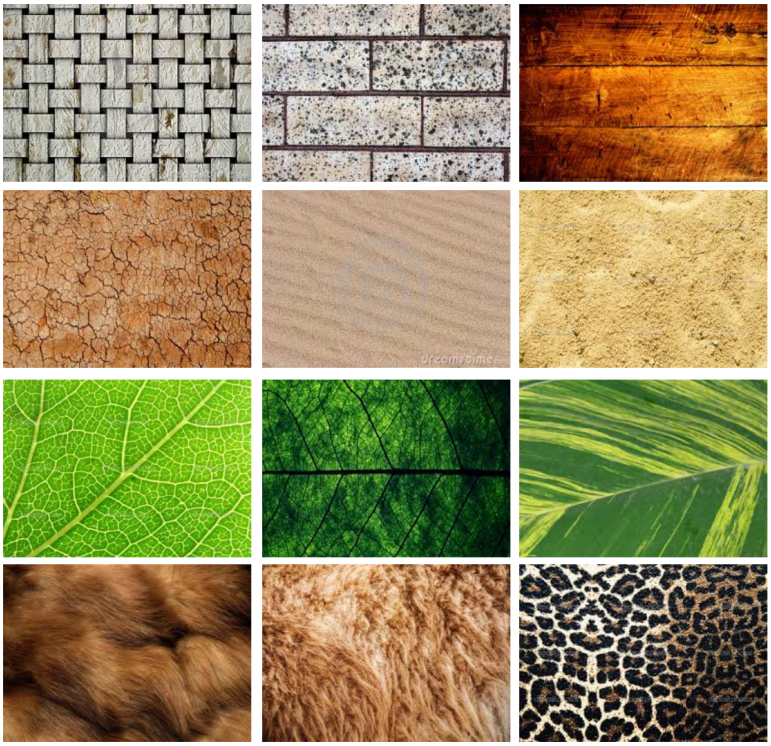

Figura 5 – Exemplos de imagens naturais com textura.

De modo geral, existem três principais abordagens utilizadas na recuperação de imagens para a criação do vetor de característica que são: a estatística, a geométrica (também chamada de estrutural ou sintática) e a espectral (denominada de métodos de processamento de sinais). A seguir são apresentados cada um deles:

- Abordagem estatística: os métodos estatísticos definem a textura em termos de distribuição espacial dos valores de tons de cinza, utilizando métodos tais como: contraste, correlação, entropia, uniformidade, densidade, aspereza, rugosidade dentre outros, para descrever suas propriedades. Estas medidas são fortemente baseadas nos aspectos de percepção humana de textura.

- Abordagem geométrica: a caracterização dos métodos geométricos são compostos de "elementos de textura" ou primitivas (por exemplo, círculos, retângulos, triângulos, etc) e, são geralmente aplicadas em texturas estritamente uniformes. Como exemplo dessas texturas, podemos citar: amostra de tecido, parede de tijolos, telhados,

dentre outros. Entretanto, esta abordagem é bastante limitada, pois a identificação automática das primitivas é considerado um problema difícil. Outra limitação da caracterização dos métodos geométricos é sua aplicação em texturas estritamente uniformes (caso muito raro em imagens reais).

- ❑ Abordagem espectral: os métodos espectrais (por exemplo, espectro de Fourier (FLORINDO; BRUNO, 2012)) utilizam a análise de frequência da imagem para classificar a textura. Descrevem a orientação de padrões periódicos ou quase periódicos em uma imagem. Esses padrões globais de textura são geralmente difíceis de se detectar com métodos espaciais devido à natureza local dessas técnicas.

Além disso, podem-se citar outras técnicas para extração de características de textura como os filtros de Gabor (RAJALAKSHMI; SUBASHINI, 2014) e as transformadas de *wavelets* (espectral) (LASMAR; BERTHOUMIEU, 2014).

Por fim, após a caracterização de uma imagem, em sistema CBIR, faz-se necessário medir a similaridade entre as representações dessas imagens. Nesse sentido, na próxima subseção, apresentamos a etapa de similaridade, bem como algumas medidas de similaridade e as formas de consultas por similaridade utilizadas na literatura.

2.2.2 Similaridade

Uma das etapas do processo de recuperação de imagens por conteúdo é a similaridade. A etapa de similaridade pode ser subdivida em duas rotinas: a medida de similaridade e a consulta por similaridade. Na primeira rotina, denominada medida de similaridade, são utilizadas funções para o cálculo de dissimilaridade entre a representação da imagem de consulta com as representações das imagens do banco, o qual será retornado um valor para cada comparação, e de acordo com o valor é possível quantificar o quão similar são as imagens comparadas. Posteriormente, na segunda rotina, definida por consulta de similaridade, deve-se escolher o operador de consulta a ser utilizado.

2.2.2.1 Medida de Similaridade

Durante o processo de recuperação de imagens baseado em conteúdo, as características extraídas são utilizadas para representar cada imagem como um ponto n-dimensional, onde n é a quantidade de características da imagem. Em seguida, para comparar a imagem de consulta com todas as imagens contidas no banco de dados, aplica-se uma medida de similaridade (função de distância) para o cálculo da dissimilaridade entre as representações das imagens, gerando uma lista ranqueada. O resultado da função de distância d entre um par de imagens é um valor real positivo. Caso as imagens sejam idênticas este valor é igual a zero, e o valor aumenta de acordo com a dissimilaridade entres

esses objetos. A seguir é apresentada a definição de uma função de distância, também chamado de espaço métrico.

Seja E um espaço vetorial e seja d uma função, em que $d : E \times E \rightarrow \mathbb{R}$ tal que

- $d(u, v) \geq 0 \ \forall \ u, v \in E$ e $d(u, v) = 0 \Leftrightarrow u = v$, $\forall u, v \in E$.
- $d(u, v) = d(v, u)$, $\forall u, v \in E$.
- $d(u, w) \leq d(u, v) + d(v, w)$, $\forall \ u, v, w \in E$.

Pode se dizer que d é uma métrica e (E, d) um espaço métrico.

Desta forma, assegurando as propriedades do espaço métrico, é possível realizar consultas por similaridade em grandes bases de imagens de modo eficiente. Entretanto, é importante salientar que, mais de uma métrica pode ser tecnicamente utilizada para medir a dissimilaridade de vetores de características, adquirido por um determinado extrator de característica, e cada métrica pode levar a um resultado diferente (BUGATTI; TRAINA; JR., 2008).

As métricas amplamente conhecidas e usadas são aquelas da família de Minkowski ou métricas L_p, que são aplicadas em domínios multidimensionais. A Eq. (1) define essas métricas pela variação do parâmetro $p \in \mathbb{R} \mid p \geq 1$, sendo que $\mathbf{x}$ e $\mathbf{y}$ são vetores em $\mathbb{R}^n$, isto é $\mathbf{x} = (x_1, ..., x_n)$ e $\mathbf{y} = (y_1, ..., y_n)$.

$$d(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{i=1}^{n} \mid x_i - y_i \mid^p} \tag{1}$$

Para $p = 1$ temos a distância de Manhattan (também conhecida como *City Block*), Euclidiana ($p = 2$) e a distância infinita (também chamada de distância de Chebychev) é o limite de (1) quando o $p \rightarrow \infty$. A Figura 6 ilustra a abrangência dessas funções em um espaço bidimensional.

Outra função utilizada para o cálculo de similaridade é a distância de Mahalanobis (MCLACHLAN, 1999), que conceitua a relação de covariância entre os atributos. Deste modo, é computada a matriz de covariância[2] V do conjunto, que é aplicada pela função de distância no cálculo de similaridade entre os vetores $\mathbf{x}$ e $\mathbf{y}$, conforme a Eq. (2).

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T V^{-1} (\mathbf{x} - \mathbf{y})} \tag{2}$$

Já a distância Canberra (LEGUAY; FRIEDMAN; CONAN, 2005) consiste em cálculos simples envolvendo a diferença absoluta dos valores das características de um vetor dividida pela soma absoluta dos mesmos. Na Eq. (3) está definida a distância Canberra, observando que se $x = y = 0$ adota-se $d(\mathbf{x}, \mathbf{y}) = 0$.

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \frac{\mid x_i - y_i \mid}{\mid x_i + y_i \mid} \tag{3}$$

[2] A matriz de covariância é uma matriz simétrica que sumariza a covariância entre N variáveis.

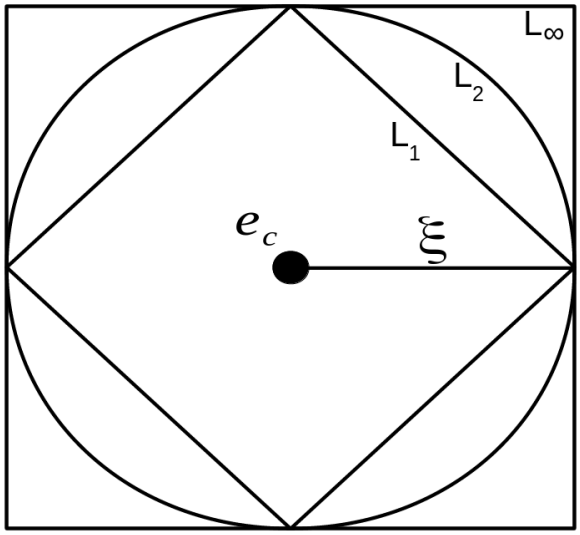

Figura 6 – Representações das formas geométricas geradas para as funções de distâncias L_1, L_2 e L_∞ para os pontos equidistantes à distância ξ a partir do elemento central e_c.

Imagem modificada de: (ROHANI; NUGOHO, 2008).

As distâncias métricas são induzidas pela norma $\|\cdot\|$ e respeitam todas propriedades métricas (simetria, positividade e desigualdade triangular). Já a medida de similaridade (também chamada aqui de função de similaridade) não precisa ser uma distância métrica, ou seja, pode não satisfazer todas as propriedades métricas. Como exemplo de medida de similaridade, pode-se citar as divergências de Bregman[3] e a Cosseno.

A similaridade Cosseno (LIU et al., 2008) compara o ângulo entre vetores, sendo que quanto menor o ângulo entre eles, maior é o grau de similaridade. A $sim(\mathbf{x}, \mathbf{y}) = 1$ quando o ângulo entre os vetores é 0°. A Eq. (4), a seguir, apresenta o cálculo da similaridade Cosseno entre os vetores $\mathbf{x}$ e $\mathbf{y}$.

$$sim(\mathbf{x}, \mathbf{y}) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|} \tag{4}$$

Existem diversas funções de similaridade na literatura, conforme explicamos acima. E, muitas vezes, a escolha da função a ser utilizada em um sistema CBIR é feita de forma aleatória, afetando o desempenho do sistema, já que cada função apresenta um comportamento diferente na comparação dos dados.

Isso porque, cada modelo de extrator de característica contém peculiaridades semânticas e uma distribuição estatística própria que devem ser levadas em consideração para a escolha da melhor medida de distância naquele determinado contexto. Por exemplo, o trabalho (LIU et al., 2008) apresenta um estudo comparativo entre quatorze medidas de dissimilaridade utilizando seis métodos de extração de características diferentes aplicados

[3] Trataremos em mais detalhes da definição da divergência de Bregman no Capítulo 4.

em um mesmo banco de dados e, para cada medida combinada com cada caracterização, obteve-se desempenho diferente.

Em sequência, apresentamos a segunda rotina inerente à etapa de similaridade na recuperação de imagens baseadas em conteúdo, a consulta por similaridade, e exemplificamos algumas das várias formas de realizá-la.

2.2.2.2 Consultas por similaridade

Uma consulta por similaridade consiste em procurar por elementos que sejam mais semelhantes a outro elemento segundo algum critério. Em geral, as consultas por similaridade realizadas em banco de dados tradicionais que manipulam dados numéricos e textuais são exatas, baseadas em operadores de ordem total ($<, \leq, >, \geq$), igualdade ($=$) e desigualdade ($\neq$). Enquanto que as consultas para recuperação de imagens baseadas em conteúdo são realizadas por meio da similaridade de características.

Os dois tipos principais de consultas por similaridade são a consulta por abrangência e a consulta aos k-vizinhos mais próximos.

A consulta por abrangência, *Range Query* (RQ), tem como objetivo encontrar todos os objetos similares até um determinado nível de similaridade, em relação com o objeto de consulta. Isto é, seja $\mathbb{O}$ o domínio dos dados, $O \subseteq \mathbb{O}$ um conjunto de objetos, $o_c \in \mathbb{O}$ um objeto de consulta, d uma função de distância definida sobre os elementos de $\mathbb{O}$ e ξ é o limiar de dissimilaridade, então uma consulta por abrangência é dada por:

$$\{o_i \in O \mid d(o_c, o_i) \leq \xi\} \tag{5}$$

A Figura 7 apresenta a ilustração bidimensional desta consulta com a função de distância Euclidiana (L_2). Os objetos de O localizado na circunferência centrado (sombreado de cinza) em o_c com raio ξ fazem parte da resposta desta consulta.

Um exemplo de consulta por abrangência em uma base de dados de imagens é "*Selecione as imagens que sejam similares à I_c até no máximo 7 unidades de distância, considerando a função de distância Euclidiana*", sendo possível escolher outra medida de distância diferente da L_2 citada no exemplo.

Existem duas variações básicas para a consulta por abrangência. A primeira é chamada de **consulta pontual**, onde o $\xi = 0$, cujo objetivo é identificar se o elemento de consulta está armazenado ou não no banco de dados, enquanto que a segunda variação é a **consulta por abrangência reversa** que tem a finalidade de procurar elementos que não estejam na área de abrangência, isto é, a resposta é formada pelos objetos $o_i \in O$ tal que $d(o_c, o_i) > \xi$.

A outra forma de consulta por similaridade comumente usada é a consulta aos k-vizinhos mais próximos. Também conhecida como *k-Nearest Neighbors Query* (k-NNq), a consulta aos k-vizinhos mais próximos, recebe como parâmetros o objeto central de consulta o_c e um número inteiro $k \geq 1$, retornando os k objetos mais próximos de o_c. Caso a cardinalidade do conjunto de dados seja menor que k, o k-NNq retorna todos os

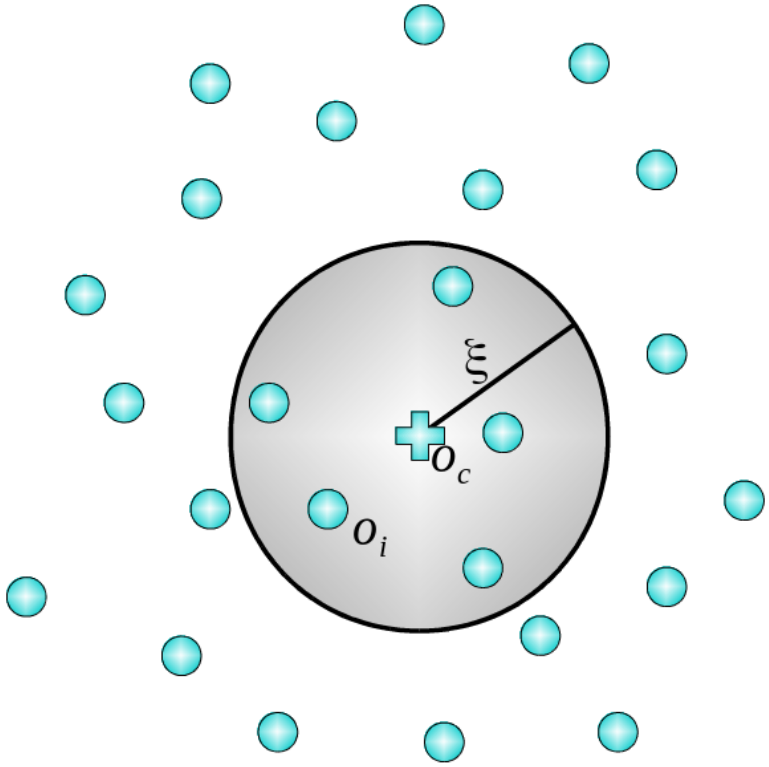

Figura 7 – Representação da consulta por abrangência no espaço bidimensional.

objetos do conjunto de dados. Além disso, pode haver dois ou mais objetos situados à mesma distância do objeto o_c que podem ser escolhidos como o k-ésimo vizinho mais próximo. Desta forma, pode-se escolher qualquer um destes elementos de forma arbitrária[4]. Formalmente, dado um domínio $\mathbb{O}$, um conjunto de elementos $O \subseteq \mathbb{O}$, um elemento de consulta $o_c \in \mathbb{O}$, uma função de distância d definida sobre $\mathbb{O}$ e um número de objetos a serem retornados igual a k, o k-NNq é dado por:

$$\text{k-NNq} = K = \{o_i \in O \mid \forall o_j \in O \setminus K, d(o_c, o_i) \leq d(o_c, o_j), \mid K \mid = k\},$$

onde $K \subseteq O$ é o conjunto resposta da consulta e $\mid K \mid$ é cardinalidade do conjunto de objetos. A Figura 8 apresenta uma 5-NNq, sendo $o_c \in \mathbb{O}$ como o elemento de consulta e neste exemplo o $o_c \notin O$. Os objetos conectados por uma linha ao o_c pertencem ao conjunto resposta. Um exemplo de consulta dos vizinhos mais próximos em uma base de dados de imagens é "*Selecione as 5 imagens mais similares a imagem o_c, considerando como função de distância L_2*".

Uma variação do k-NNq é a **consulta aos *k*-vizinhos mais distantes** (*k-Farthest Neighbors query* – *k*-FNq), que em vez de procurar os vizinhos mais próximos, faz a busca pelos k objetos mais dissimilares ao objeto o_c.

[4] Neste trabalho foi adotado a escolha arbitraria quando ocorre o empate, como a maioria dos trabalhos encontrados na literatura.

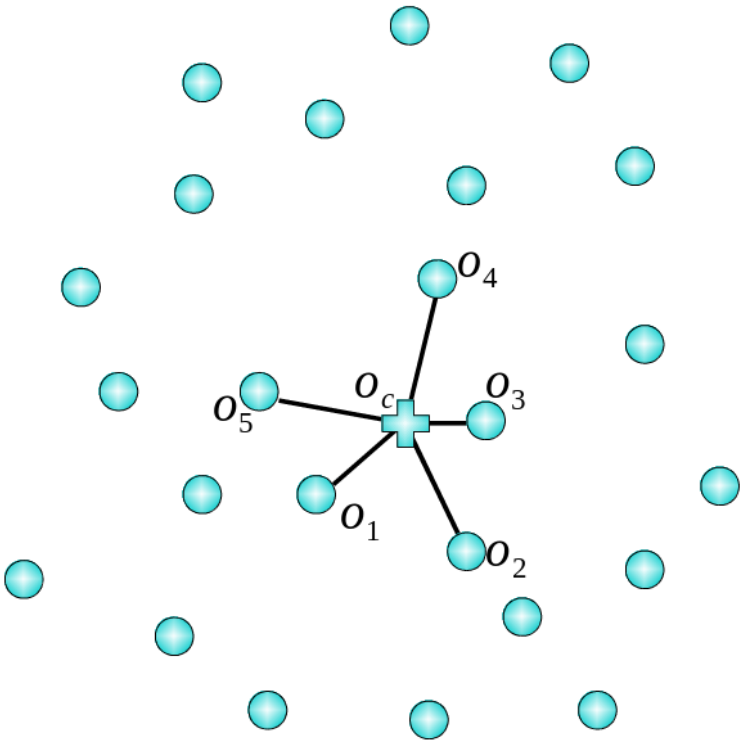

Figura 8 – Consulta aos k-vizinhos mais próximos com $k = 5$ sobre o objeto o_c no espaço bi-dimensional com a função de distância Euclidiana.

Após definida a caracterização para representar a imagem, a função de distância e o tipo de consulta para realizar a etapa de similaridade no sistema CBIR, torna-se extremamente relevante analisar o desempenho do sistema para averiguar a qualidade dos resultados.

Em sequência, são apresentados alguns métodos para avaliar o sistema de recuperação de imagem baseada em conteúdo.

2.3 Métodos de Avaliação dos sistemas de recuperação

Compreendendo todo o processo de recuperação de imagens baseado em conteúdo e suas técnicas, podemos explanar sobre alguns métodos conhecidos para mensurar o desempenho de um sistema CBIR. Detalhamos aqui os seguintes métodos de avaliação de desempenho de sistemas: as medidas de precisão e revocação, precisão em k, *Mean Average Precision* (MAP) e o (*Discounted Cumulative Gain* (DCG)).

2.3.1 Precisão x Revocação

As duas medidas mais eficazes e frequentemente utilizadas no contexto de recuperação de imagens são a precisão e a revocação. Estas medidas são, primeiramente, definidas por um caso simples, por exemplo, sistema de recuperação de informação (*Information Retrieval* (IR)) que tem como objetivo retornar um conjunto de documentos de acordo com a consulta (*query*) (MANNING; RAGHAVAN; SCHÜTZE, 2008).

A precisão (P) é a fração de documentos recuperados que são relevantes, já a revocação (R) é a fração de documentos relevantes que são recuperados (WEN; ZHANG; RAMAMOHANARAO, 2014). Estas noções podem ficar mais claras examinando a seguinte tabela de contingência (veja a Tabela 1):

Tabela 1 – Tabela de contingência (modificado de (MANNING; RAGHAVAN; SCHÜTZE, 2008)).

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

A precisão e revocação são formuladas da seguinte maneira:

$$P = \frac{vp}{(vp + fp)} \tag{6}$$

$$R = \frac{vp}{(vp + fn)} \tag{7}$$

As medidas de precisão e revocação concentram-se na avaliação do retorno de verdadeiros positivos, perguntando qual a porcentagem dos documentos relevantes que são encontrados e quantos falsos positivos também foram retornados. Em um bom sistema, a precisão geralmente diminui à medida que o número de documentos recuperados aumenta. Em geral, deve-se tolerar apenas uma certa quantidade de revocação enquanto admite apenas uma certa porcentagem de falsos positivos.

A Figura 9 mostra um exemplo de gráfico de precisão e revocação, no qual duas curvas são apresentadas, X e Y. De acordo com a Figura 9 as curvas apresentam comportamentos diferentes, ou seja, os algoritmos aplicados para a recuperação são distintos. Analisando a curva X, nota-se que a mesma contém valores altos de precisão para níveis de revocação baixos, significando que a busca realizada pelo usuário retorna as imagens relevantes nas primeiras posições, o que pode ser interessante quando apenas as 20 ou 30 primeiras imagens são importantes. Enquanto que a curva Y apresenta maior precisão que a curva X para níveis de alta revocação, este comportamento é ideal para um usuário que deseje garantir que todas as imagens relevantes foram recuperadas de fato.

Em casos em que para o usuário o importante é a quantidade de resultados bons que serão exibidos na primeira página ou nas três primeiras páginas, particularmente nas pesquisas na web, é interessante utilizar a precisão em uma posição fixa, chamada de precisão em k.

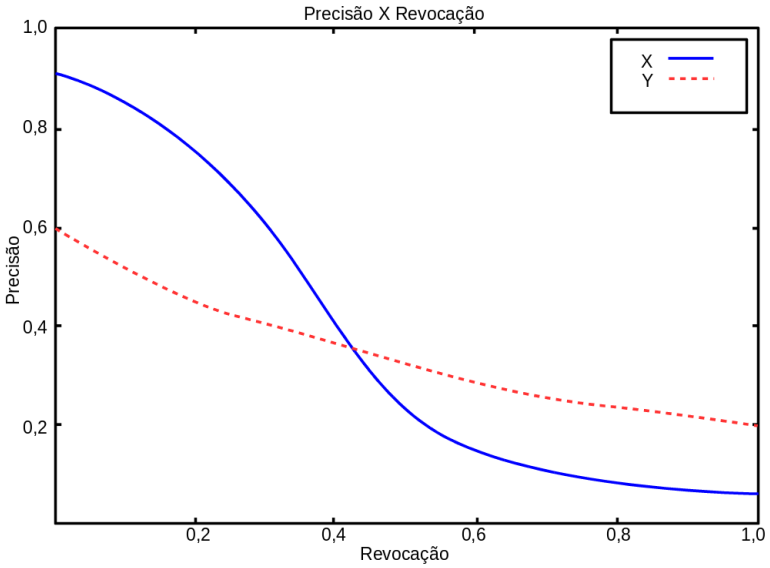

Figura 9 – Exemplo de gráfico de precisão e revocação.

2.3.2 Precisão em *k*

A precisão em k (*precision at k*) é um fator de medida em todos os níveis de revocação, e que objetiva medir a precisão fixa em baixos níveis de resultados recuperados, tal como 10 ou 30 documentos. Isto é chamado como "*precision at k*", onde o k é a posição fixa da precisão, por exemplo "precisão em 10". A vantagem é que não é necessário avaliar qualquer tamanho do conjunto de documentos relevantes. E as desvantagens são mínimas, em que as medidas de avaliação normalmente utilizadas, e que não têm uma boa medida (ou seja, número pequeno de documentos relevantes para uma consulta), tem uma forte influência na precisão em k.

Apresentamos a seguir uma outra medida que tem mostrado uma boa descriminação e estabilidade, chamada de *Mean Average Precision* (MAP), a seguir.

2.3.3 *Mean Average Precision*

Entre as medidas de avaliação, o *Mean Average Precision* (MAP) é o que tem demonstrado uma boa discriminação e estabilidade, fornecendo uma medida de um único valor de qualidade entre os diferentes níveis de revocação (MANNING; RAGHAVAN; SCHÜTZE, 2008). O MAP é a média do *Average Precison* (AP), e a AP pode ser definida como a média dos valores das precisões obtidas pelo conjunto de *top* k documentos existentes após cada documento relevante recuperado, assim, calcula-se a média dos valores sobre as informações das precisões. Isto é, para uma única consulta $q_j \in Q$, AP é a média das precisões computadas no ponto de cada item recuperado corretamente $\{d_1, ..., d_{m_j}\}$ na

lista ranqueada, e este valor é então calculado sobre o conjunto de consultas Q:

$$MAP(Q) = \frac{1}{\mid Q \mid} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} precisão(R_{jk}) \tag{8}$$

Onde R_{jk} é o conjunto de resultados ranqueados, iniciando dos melhores resultados até chegar ao item d_k. Quando um documento relevante não é recuperado de todos[5], o valor da precisão na Eq. (8) é feita para ser 0.

Utilizando o MAP, os níveis de revocação fixos não são escolhidos e não possuem interpolação. O valor MAP para uma coleção de teste é a média aritmética dos valores das precisões médias 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 consultas.

Por fim, apresentamos, a seguir, uma medida de avaliação que realiza uma ponderação nos resultados, sendo que os resultados corretos que estão nas primeiras respostas têm um peso maior do que as que estão nas últimas posições das respostas, esta medida é chamada de *Discounted Cumulative Gain* (DCG).

2.3.4 *Discounted Cumulative Gain*

A avaliação de desempenho *Discounted Cumulative Gain* (DCG) foi proposta por Järvelin e Kekäläinen (JÄRVELIN; KEKÄLÄINEN, 2002), e é considerada como uma estatística que pondera resultados corretos. Os resultados localizados nas posições na frente de uma lista, têm um peso maior do que os resultados corretos mais ao final da lista de classificação, supondo que o usuário não considerará os elementos próximos ao fim da lista.

(DUPRET; PIWOWARSKI, 2013) relatam que existem duas interpretações para essa métrica, que são a utilitário e a probabilística. Olhando do ponto de vista utilitário, considera-se que a utilidade de um documento para um usuário diminui quando o documento tem um ranqueamento baixo. Já para o probabilístico, considera-se que todos os documentos não são examinados com a mesma probabilidade, isto é motivado pelo fato que a escolha do documento tem uma probabilidade de acordo com sua posição no *rank*, ou seja, se o documento está nas últimas posições, a probabilidade de ser escolhido pelo usuário é menor do que o documento que está nas primeiras posições.

Especificadamente, a lista de classificação R é convertida para uma lista G, em que os elementos G_i têm valor 1 se os elementos R_i estão na classe correta e o valor 0 caso

[5] Um sistema não pode ordenar todos os documentos em uma coleção como resposta para uma consulta, ou pelos menos pode ser baseado no envio de apenas os *top k* resultados para cada informação da precisão.

contrário. O DCG_k é então definido como a seguir (SHILANE et al., 2004):

$$DCG_k = G_1 + \sum_{i=2}^{k} \frac{G_i}{\log_2(i)} \tag{9}$$

onde o k é a posição do *ranking* da lista G.

Para que seja possível a comparação do desempenho de diferentes algoritmos, os valores dos DCGs obtidos de cada algoritmo devem ser normalizados (MANNING; RAGHAVAN; SCHÜTZE, 2008). Então, o resultado obtido da Eq. (9) é dividido pelo DCG_k máximo possível que corresponde ao ranqueamento perfeito, que será chamado de IDCG_k – Ideal DCG_k (por exemplo, seria os k primeiros elementos que estão todos classificados na classe correta). A Eq. (10) mostra a normalização do DCG (nDCG).

$$nDCG_k = \frac{DCG_k}{IDCG_k}, \quad IDCG_k = 1 + \sum_{i=2}^{k} \frac{1}{log_2(i)}, \quad k \leq \mid C \mid \tag{10}$$

onde $\mid C \mid$ é o número de elementos relevantes.

Os valores de nDCG_k variam no intervalo de [0,1]. Os resultados do nDCG de todas as consultas podem ser totalizados em uma média aritmética para que seja usado como valor do desempenho do algoritmo utilizado e quanto maior este valor, melhor é o resultado.

Por fim, foram apresentados neste capítulo alguns aspectos importantes a serem considerados no processo de recuperação de imagem baseado em conteúdo como: extração de características, medidas de similaridade, consultas por similaridade e métodos de avaliação de sistemas CBIR para avaliar a qualidade dos resultados obtidos.

No próximo capítulo são descritas algumas abordagens para caracterizar um conjunto de imagens com intuito de otimizar a recuperação.

Capítulo 3

Abordagens que utilizam *Bag-of-Visual-Words*

Além das formas descritas neste trabalho para realizar a caracterização, existem na literatura abordagens para caracterizar um conjunto de imagens com objetivo de otimizar a recuperação e/ou classificação do conjunto, a fim de reduzir o *gap* semântico entre características de baixo nível e o conteúdo visual da imagem, por exemplo, a abordagem *Bag-of-Visual-Words* (BoVW). A BoVW pode ser combinada com diversas técnicas, inclusive a de atenção visual (na etapa de caracterização do conjunto de imagens) ou as pirâmides espacias (fase de geração dos histogramas). Dentre as abordagens que utilizam a atenção visual podem-se citar o FISM e o BSM. O BoVW-SPM é a combinação do BoVW com pirâmides espacias (*Spatial Pyramids Matching* – SPM) para gerar e quantificar os histogramas de acordo com a divisão da imagem em sub-regiões com diferentes níveis de relevância. Por fim, são concatenados os histogramas correspondentes a cada sub-regiões da imagem, criando um histograma que irá representar a imagem. Assim, a seguir, apresentamos detalhes da técnica BoVW, e suas respectivas abordagens, utilizadas para a caracterização visual de imagens nos sistemas CBIR.

3.1 *Bag-of-Visual-Words*

A técnica BoVW[1] foi originada da técnica *Bag-of-Words* (BoW), utilizada na área de recuperação de informação (RENALS et al., 2000) com intuito de recuperar textos.

A abordagem BoW tem a finalidade de representar um documento textual como um conjunto de palavras, que faz parte de um vocabulário fixo, obtido por meio de uma base de documentos, ignorando qualquer estrutura inerente ao documento (VALLE; CORD, 2009). Sua função estima a probabilidade de uma palavra estar contida em um determinado contexto.

[1] Também chamado de *bag-of-keypoints*, *bag-of-features* ou *bag-of-visual-features* para se referir ao mesmo método.

O modelo BoW obteve um grande sucesso em sistemas de recuperação de documentos, e no trabalho de (ZHU; RAO; ZHANG, 2002) adaptou-se esta abordagem para categorização visual, criando uma quantização de vetor de pequenas janelas de imagens quadradas, que foram denominadas de blocos-chave. Esta nova abordagem foi denominada de BoVW, que é uma técnica de representação das características visuais de um determinado conjunto de imagens com objetivo de otimizar a recuperação e/ou classificação do conjunto a fim de reduzir a diferença semântica entre as características de baixo nível e o conteúdo visual da imagem. A Figura 10, a seguir, demonstra o fluxograma de funcionamento da BoVW que pode ser combinado com atenção visual e o SPM.

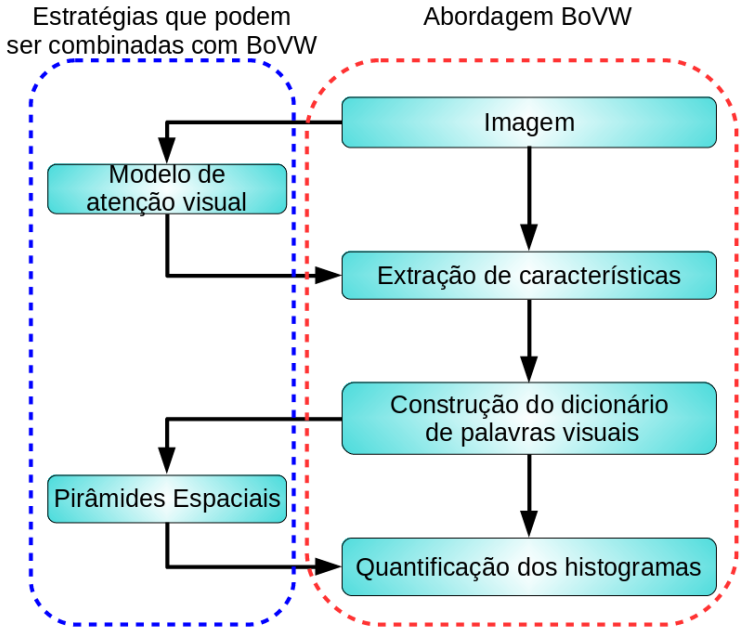

Figura 10 – Fluxograma da abordagem BoVW combinada com os métodos de atenção visual e pirâmides espaciais.

O uso da técnica BoVW tem se tornado importante em atividades de pesquisa na área de visão computacional. Como exemplo, podem-se citar os trabalhos de Dong (DONG; GUO; FU, 2014), Pedrosa (PEDROSA; TRAINA; JR., 2014), Godil (GODIL; LIAN; WAGAN, 2013), Grzeszick (GRZESZICK; ROTHACKER; FINK, 2013), Amato (AMATO; FALCHI; GENNARO, 2013), Wang (WANG, 2012; WANG et al., 2011), Soares (SOARES; SILVA; GULIATO, 2012) e Liu (LIU et al., 2011).

O funcionamento do BoVW pode ser dividida em três principais etapas: extração de característica, construção do dicionário de palavras visuais e quantificação dos histogramas utilizando o vocabulário visual. As etapas são descritas a seguir:

- A **extração de característica** consiste em representar pontos locais de interesse em um conjunto de imagens, utilizado para o aprendizado por meio de alguma técnica, como por exemplo, detector de ponto de interesse (SIFT) (LOWE,

2004; NGUYEN et al., 2015), *Principal Component Analysis*-SIFT (PCA-SIFT) (ZICKLER; EFROS, 2007), *Speeded Up Robust Features* (SURF) (MENDOZA-MARTINEZ; ORTEGA; ARREGUIN, 2014), amostragem aleatória (ULLMAN; VIDAL-NAQUET; SALI, 2002). O conjunto de imagens será representada por um conjunto de descritores, que são vetores de dimensão elevada, tais como os descritores SIFT. Estes vetores são denominados de características e são utilizados para construir o dicionário de palavras visuais. É de suma importância que os descritores sejam invariantes às condições de transformações na imagem como rotação, translação, iluminação e oclusões parciais.

- **Construção do dicionário de palavras visuais** é a etapa posterior da extração de características das imagens. Geralmente são utilizadas técnicas de agrupamento, tal como o *k-means* (LIBERTY; SRIHARSHA; SVIRIDENKO, 2014; ELKAN, 2003; FORGY, 1965), para gerar os vocabulários. Nesse momento, os centróides de cada agrupamento são considerados como sendo uma palavra visual e o conjunto dessas palavras formam o vocabulário (também chamado de *codebook* ou dicionário).

- Para a **construção dos histogramas de palavras visuais** todas as características de cada imagem são mapeados para a palavra visual mais próxima, obtendo assim um histograma de palavras visuais associadas a cada imagem do banco de dados. O histograma resultante é conhecido como BoVW e sua dimensão está associada 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.

Após a quantificação dos histogramas pode-se fazer a busca por similaridade. O cálculo da similaridade é realizado entre os histogramas das imagens da base de dados e o histograma da imagem de consulta utilizando algum operador de similaridade. Quanto menor a distância entre os histogramas, mais similares eles são.

A abordagem BoVW tem demonstrado bons resultados em diferentes aplicações de identificação de objetos e cenas. Entretanto, a simplicidade da representação é, ao mesmo tempo, seu ponto forte e seu ponto fraco, pois informações de espacialidade ou dependência das palavras visuais são ignoradas.

A seguir é apresentado um descritor que é bastante utilizado na abordagem BoVW denominado de *Scale-Invariant Feature Transform* (SIFT).

3.1.1 *Scale-Invariant Feature Transform*

A SIFT foi desenvolvida em 1999 por David G. Lowe, professor do departamento de Ciência da computação da *University of British Columbia*. Inicialmente o descritor SIFT tem sido proposto para possibilitar eficientes tarefas de reconhecimento de objetos

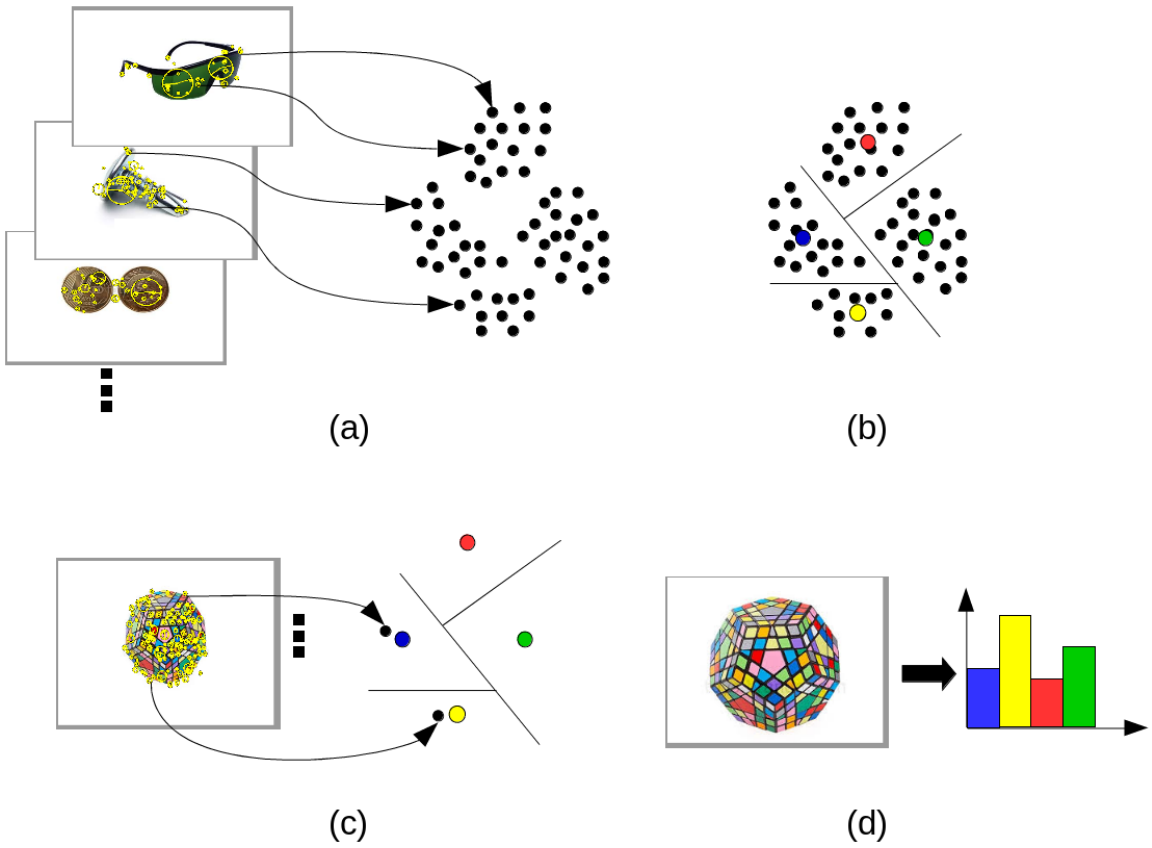

Figura 11 – Visão geral do *Bag of Visual Words*. a) Uma grande amostra de características locais são extraídos a partir de um conjunto de imagens. Os círculos amarelos nas imagens representam as características locais e os círculos pretos denotam pontos em algum espaço de características dos pontos chaves, por exemplo o SIFT. b) Realiza a clusterização dos pontos chaves para gerar as palavras visuais (representandos pelos círculos coloridos) e, por fim formar o vocabulário. c) Dada uma nova imagem, são extraídas suas características e mapeadas para a palavra visual mais próximas. d) E finalmente é criado um histograma de palavras visuais para cada imagem.

(LOWE, 1999; LOWE, 2004). Em recentes trabalhos, esta técnica tem sido explorada no método *Bag of Visual Words* (SIVIC; ZISSERMAN, 2003) (apresentado na seção 3.1). Devido ao poder discriminativo e estabilidade do descritor SIFT, tornou-se bastante utilizado como descritor de pontos-chave em uma infinidade de tarefas. Algumas aplicabilidades do descritor SIFT são: reconhecer objetos em imagens, modelagem 3D, rastreamento, reconhecimento de gestos humanos, *tracking* de vídeo, dentre outros (LOWE, 1999).

As características obtidas pelo descritor SIFT são bem localizadas nos domínios de frequência e do espaço, reduzindo assim a probabilidade de não haver correspondência das características por oclusão ou ruído. Essas características são bem distintas, possibilitando que uma simples característica seja corretamente correspondida com alta probabilidade diante de um grande banco de dados de características (LOWE, 2004).

O funcionamento do descritor SIFT segundo (LOWE, 2004) é dividido em quatro etapas principais, que são a detecção de extremos, localização de pontos-chave, definição

da orientação e descrição dos pontos-chave. A seguir são descritas essas etapas.

- **Detecção de extremos**: neste primeiro estágio é realizada a procura por todas as escalas e posições de uma imagem. Para isso é usada uma função conhecida como Diferença Gaussiana (*Difference of Gaussian* (DoG)) (RAFIEE; DLAY; WOO, 2013) para detectar os extremos (máximos e mínimos) da imagem, com o intuito de identificar os potenciais pontos de interesse, que são invariantes à orientação e escala.

- **Localização dos pontos-chave**: esta etapa tem como definir quais pontos de interesse serão candidatos para serem descritos na última etapa. Para cada candidato é determinada a sua posição, escala e razão das curvaturas principais. Esta razão tem a finalidade de auxiliar na rejeição dos pontos que possuem baixo contraste ou que estão localizados em bordas não definidas. Para localizar a posição e escala para os pontos candidatos, é ajustada uma função quadrática 3D ao ponto de amostragem local de modo a determinar uma localização interpolada máxima. Isto é feito por meio de uma expansão de Taylor da função DoG aplicada à imagem. Deste modo é feita a seleção dos pontos chaves de acordo com suas medidas de estabilidade.

- **Definição da orientação**: para cada ponto-chave são atribuídas uma ou mais orientações para cada ponto-chave localizado, baseadas em direções do gradiente. Para calcular a magnitude e orientação do gradiente utilizam-se as diferenças de *pixels*, e, em seguida, é construído o histograma de orientações para os *pixels* em torno do ponto-chave. As direções dominantes dos gradientes locais são representados pelos picos nos histogramas, permitindo assim definir a orientação.

- **Descrição dos pontos-chave**: uma região de 16×16 *pixels*, localizada no ponto-chave central é subdividido em 4×4 sub-regiões. Essas 16 sub-regiões são rotacionadas em relação à orientação canônica computadas para o ponto-chave. Para cada sub-região, um histograma com 8 *bins* de orientação são computados. O valor da magnitude para todos os gradientes dentro da região são ponderados por uma janela Gaussiana e acumulado nos histogramas de orientação. Os 8 *bins* de todos os 16 histogramas são concatenados formando um vetor de 128 dimensões, o qual em seguida é normalizado para ter invariância à iluminação, assim representando o descritor SIFT.

A Figura 12 apresenta o resultado da aplicação do SIFT na detecção de pontos de interesse de uma imagem. A dimensão de cada circunferência corresponde à escala do respectivo ponto-chave, e os raios definem a sua orientação. Dependendo da simetria do ponto-chave, a determinação da orientação pode ser ambígua, fazendo que tenha mais do que uma possível orientação.

Para cada imagem são construídos diversos descritores, cada um referente a um ponto-chave. Quando é aplicado o descritor SIFT em uma imagem, o resultado é um conjunto

de descritores. Várias extensões do SIFT têm sido propostas na literatura, por exemplo o PCA-SIFT (ZICKLER; EFROS, 2007) que aplica o PCA em *patches* de gradientes normalizados para reduzir o tamanho do descritor SIFT original. A *Rotation-Invariant Feature Transform* (RIFT) (LAZEBNIK; SCHMID; PONCE, 2005) que divide cada *patch* da imagem dentro de anéis concêntricos de largura igual, para superar o problema de estimativa da orientação dominante do gradiente exigido pelo SIFT. O *Rank*-SIFT (LI et al., 2011) que define cada *bin* do histograma para sua classificação em uma matriz ordenada de *bins*.

Quando se aplica o descritor SIFT para tarefas como classificação de objetos ou de cenários, sobre uma grade densa no domínio da imagem é chamado de *Dense Scale-Invariant Feature Transform* (D-SIFT) (VEDALDI; FULKERSON, 2010). Utilizando o D-SIFT obtém-se um descritor com mais características de cada localização e escala em uma imagem, fazendo com que sua complexidade computacional aumente, comparada ao SIFT. Caso contrário, o SIFT é aplicado nos pontos de interesses espasos no domínio da imagens, sendo denominado de *Sparse Scale-Invariant Feature Transform* (S-SIFT).

Figura 12 – Exemplos de pontos-chave detectados pelo SIFT.

Na seção seguinte será abordado alguns descritores que combinam o BoVW com atenção visual.

3.2 Descritores de características considerando a percepção visual humana

A abordagem BoVW é uma estratégia de recuperação de imagens bastante utilizada para o reconhecimento de objetos, sendo bastante comum o uso do descritor SIFT para etapa de extração de características. Entretanto, a BoVW não utiliza a percepção visual

humana para saber o que é relevante ou não na imagem. Por isso, nesta Seção conceituamos o que é a atenção visual e, em seguida, apresentamos duas abordagens: a BSM e a FISM, que utilizam a atenção visual baseada na extração de mapas de saliência, aplicando estas técnicas a uma imagem, conseguem identificar o que é relevante de acordo com a percepção humana.

3.2.1 Atenção Visual

A todo momento, os olhos humanos se deparam com uma grande carga de estímulos visuais. No entanto, é impossível processar toda a informação que chega aos olhos de uma só vez (TSOTSOS, 1990). Para contornar essa situação, o sistema visual humano seleciona e processa rapidamente apenas as regiões de interesse em uma determinada imagem visual, e este mecanismo é nomeada de atenção visual. A seleção das regiões de interesse é importante para reduzir a quantidade de informações a serem processadas (FISCHER; WEBER, 1993).

Na área de processamento de imagens, os sistemas computacionais têm uma grande dificuldade com o grande volume de informações a serem processadas. Assim, os modelos de atenção visual como ferramentas computacionais tentam imitar de forma qualitativa o comportamento do sistema visual humano. De modo que a atenção visual torna-se uma alternativa interessante para auxiliar na redução da quantidade de dados processados, aumentando a eficiência do sistema e permitindo que os recursos computacionais sejam utilizados para processar apenas regiões de interesse na cena.

Em termos computacionais, a atenção visual funciona da seguinte maneira: primeiro calcula-se um conjunto de características em paralelo, para então depois combiná-las em uma representação chamada de mapa de saliência. Em geral, essas características são intensidade, orientação, cor, movimento, faces, gestos, dentre outras (FRINTROP; ROME; CHRISTENSEN, 2010).

Existem vários modelos computacionais de atenções visuais disponíveis na literatura como (NIEBUR; KOCH, 1996), (ITTI; KOCH; NIEBUR, 1998), (NIEBUR; KOCH, 1998), (ITTI; KOCH, 2001), (HAREL; KOCH; PERONA, 2007) e (RAJASHEKAR et al., 2008). De forma geral, a atenção visual está dividida em duas principais abordagens, *top-down* e *bottom-up*.

Os modelos *top-down* levam em consideração algum conhecimento derivado a partir de experiências anteriores ou gostos pessoais para determinar as regiões de interesse na imagem, ou seja, utilizam características de alto nível das imagens como modelos geométricos, estatísticos, dentre outros, para encontrar as regiões de maior interesse da cena. Por exemplo, se uma pessoa está procurando por um objeto com uma forma geométrica específica (por exemplo um caderno), aspectos de mais alto nível podem guiar o processo seletivo de atenção durante a busca e ignorar outras características visuais irrelevantes ou que não compõem o objeto em questão.

Enquanto que os modelos *bottom-up* se baseiam no princípio que, a atenção é atraída para locais específicos da cena (regiões salientes), ou seja, estas regiões não são diferentes o suficiente do ambiente que os rodeiam, considerando características de baixo nível das imagens (como cor, intensidade e orientação), sem qualquer informação contextual para definir a atenção visual. Por exemplo, considere uma cena a ser analisada composta por vários objetos de cores azuis e um único objeto de cor vermelha, tem-se a sensação de que o objeto de destaque é o de cor vermelha. Este fenômeno é resultante do alto contraste no atributo cor entre os objetos, pelo fato do objeto vermelho se destacar, o torna um melhor candidato durante o processo de competição por atenção.

Enfatizaremos aqui a abordagem *bottom-up*, empregando um modelo de mapa de saliência proposto por Itti (ITTI; KOCH, 2001). O mapa de saliência apresentado por Itti (ITTI; KOCH; NIEBUR, 1998; ITTI; KOCH, 2001) simula as propriedades de baixo nível do sistema visual humano e se baseia na extração de mapas de saliência. O modelo proposto é gerado a partir da integração das seguintes características primitivas como cor, intensidades e orientação.

O mapa de saliência (MS) é definido por um mapa de duas dimensões responsáveis por codificar as saliências sobre todos os pontos da cena visual, baseando-se na ideia de que a atenção é direcionada para a diferença de contraste local de atributos visuais (ITTI; KOCH; NIEBUR, 1998). O modelo de Itti é dividido nas seguintes etapas: filtragem linear, diferença centro-vizinhança e normalização, combinações variando a escala e normalização e por fim a combinação linear. A Figura 13 apresenta a arquitetura geral do modelo.

Inicialmente, as características visuais da imagem de entrada passa pelo módulo de filtragem linear, em seguida são extraídas a cor, intensidade e orientação. Com a extração dos canais r, g e b (*red green blue*), a intensidade da imagem I é obtida como $I = (r+g+b)/3$, que também representa/define a imagem em tons de cinzas. Para extrair os quatro canais (R-vermelho, G-verde, B-azul e Y para amarelo) são criados desta forma: $R = r-(g+b)/2$, $G = g-(r+b)/2$, $B = b-(r+g)/2$ e $Y = (r+g)/2- \mid r-g \mid /2-b$. A imagem I e os canais R, G, B e Y são utilizadas para criar a Pirâmide Gaussiana (GREENSPAN et al., 1994) $I(\sigma)$, $R(\sigma)$, $G(\sigma)$, $B(\sigma)$ e $Y(\sigma)$, onde $\sigma \in [0..8]$ é a escala.

Os mapas de características são obtidos por meio da diferença entre canais de cores em diferentes escalas, sendo que este processo é denominado de centro-vizinhança definido por $\ominus$. O $\ominus$ entre um "centro" (*pixel*) $c \in \{2,3,4\}$ e sua vizinhança corresponde a escala $s = c+\delta$, em que $\delta \in \{3,4\}$. A diferença *across-scale* entre dois mapas, denotado por "$\ominus$", é obtido pela interpolação da escala mais fina e a subtração ponto-a-ponto. O primeiro conjunto de mapas de características obtido é a intensidade do contraste que é computado 6 mapas $\mathcal{I}(c,s)$:

$$\mathcal{I}(c,s) = \mid I(c) \ominus I(s) \mid \tag{11}$$

O segundo conjunto de mapas (que contém um total de 12) são construídos para os

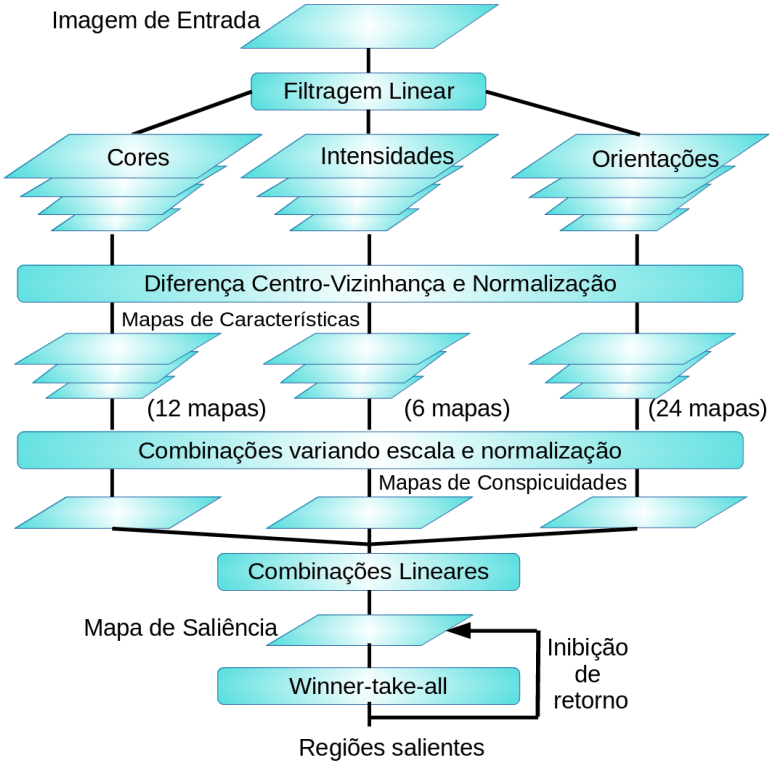

Figura 13 – Modelo geral de atenção visual baseado em mapa de saliência. Imagem modificada de (ITTI; KOCH; NIEBUR, 1998).

canais de cor. Os mapas $\mathcal{RG}(c,s)$ são criados no modelo para representar simultaneamente vermelho/verde e verde/vermelho (mostrado na Eq. (12)), e $\mathcal{BY}(c,s)$ para azul/amarelo e amarelo/azul como apresentado na Eq. (13).

$$\mathcal{RG}(c,s) = \mid (R(c) - G(c)) \ominus (G(s) - R(s)) \mid \tag{12}$$

$$\mathcal{BY}(c,s) = \mid (B(c) - Y(c)) \ominus (Y(s) - B(s)) \mid \tag{13}$$

Os 24 mapas de orientação são obtidos de I utilizando as pirâmides de Gabor $O(\sigma,\theta)$ onde o σ representa a escala e $\theta \in \{0°, 45°, 90°, 135°\}$ é a orientação preferida (GREENSPAN et al., 1994). A Eq. (14) apresenta a fórmula para extração dos mapas de orientação.

$$\mathcal{O}(c,s,\theta) = \mid O(c,\theta) \ominus O(s,\theta) \mid \tag{14}$$

No total são computados 42 mapas de características, seis para intensidade, 12 para cor e 24 para orientação. Combinando uma pequena quantidade de mapas fazem com que os objetos sejam vistos mais nitidamente. Para adquirir uma saliência menor dos objetos, deve-se fazer uma combinação maior destes mapas de características.

O propósito do mapa de saliência é representar a conspicuidades – ou "saliência" – em qualquer localização no campo visual por uma quantidade escalar e guiar a localização baseado na distribuição espacial da saliência.

Os mapas de características são combinados em três "mapas de conspicuidades", $\overline{\mathcal{I}}$ para intensidade como mostrado na Eq. (15), $\overline{\mathcal{C}}$ para cor (Eq. (16)), e orientação $\overline{\mathcal{O}}$ (Eq. (17)), na escala $\sigma = 4$ para o mapa de saliência. Onde o $\mathcal{N}(.)$ é operador normalizador, que promove globalmente os mapas em um número pequeno de fortes picos de atividade (localizações de conspicuidades) que estão presente, enquanto que os suprime globalmente os mapas que contém numerosas respostas dos picos comparáveis. O operador $\bigoplus$ representa a adição em escala, que consiste de redução de cada mapa em escala 4 e adição ponto-a-ponto:

$$\overline{\mathcal{I}} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{I}(c,s)) \tag{15}$$

$$\overline{\mathcal{C}} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [\mathcal{N}(\mathcal{RG}(c,s)) + \mathcal{N}(\mathcal{BY}(c,s))] \tag{16}$$

$$\overline{\mathcal{O}} = \sum_{\theta \in \{0°,45°,90°,135°\}} \mathcal{N}\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{O}(c,s,\theta)) \right) \tag{17}$$

Os três mapas de conspicuidades são somados e normalizados para a saída final do mapa de saliência, representada por $\mathcal{S}$:

$$\mathcal{S} = \frac{1}{3}(\mathcal{N}(\overline{\mathcal{I}}) + \mathcal{N}(\overline{\mathcal{C}}) + \mathcal{N}(\overline{\mathcal{O}})) \tag{18}$$

O mapa de saliência obtido é uma imagem em tons de cinza, a qual define as regiões mais salientes na imagem, ou seja, aonde a atenção visual será direcionada. Uma vez gerado $\mathcal{S}$, alimenta uma rede neural *Winner-Take-All* (KOCH; ULLMAN, 1985) que garante a manutenção das regiões mais importantes, enquanto que outras regiões são inibidas. A Figura 14 apresenta um exemplo que contém uma imagem e o seu respectivo mapa de saliência proposto pelo modelo desenvolvido pelo Itti.

Figura 14 – Mapa de saliência obtido através do modelo proposto por Itti.

A seção posterior apresenta duas abordagens que utilizam o BoVW combinado com o modelo computacional de atenção visual proposto por (ITTI; KOCH; NIEBUR, 1998; ITTI; KOCH, 2001), denominados de *Binary Image Descriptor Saliency Map* (BSM) e *Fuzzy Descriptor Image Saliency Map* (FISM).

3.2.2 *Binary Image Descriptor Saliency Map* (BSM)

Os mapas de saliência podem ser vistos como uma função de pertinência nebulosa, e os valores altos representam pontos na imagem que devem ter uma atenção visual maior. Estes pontos que retêm maior atenção visual serão referidos como regiões de interesse (*foreground*), enquanto que os pontos de menor interesse serão considerados como o fundo da imagem (*background*).

Os trabalhos de (NAKAMOTO; TORIU, 2011; SOARES; SILVA; GULIATO, 2012) apresentam uma forma de separar *foreground* de *background* utilizando mapas de saliência por separação binária. Este método funciona da seguinte maneira: a área onde a saliência é maior pode ser extraída aplicando uma operação de limiar (*threshold*) no mapa de saliência. Esta técnica de aplicar um determinado limiar no mapa de saliência para separar o *foreground* do *background* é denominada de mapa de saliência binária.

O método mapa de saliência binária utiliza o mapa de saliência (MS) para definir os *pixels* que pertencem ao conjunto *foreground* ou *background* da seguinte maneira (SOARES; SILVA; GULIATO, 2012): após obter o MS (utilizando o modelo de Itti (ITTI; KOCH; NIEBUR, 1998) ou algum outro) de uma determinada imagem I, os valores do MS são normalizados para o intervalo [0, 1]. Em seguida é realizada a comparação utilizando o *threshold* t com os valores do MS, se o $MS(i, j) > t$ o *pixel* fará parte do *foreground*, caso contrário $MS(i, j) \leq t$, então o *pixel* pertencerá ao conjunto do *background*. A Figura 15 apresenta uma imagem e seu respectivo mapa de saliência criada a partir do modelo de Itti (ITTI; KOCH; NIEBUR, 1998).

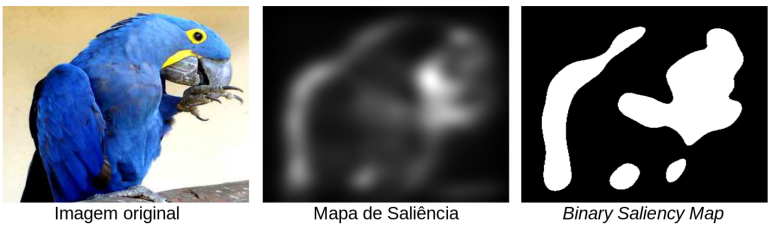

Figura 15 – Apresenta a imagem original, o mapa de saliência criada pelo modelo de Itti (ITTI; KOCH; NIEBUR, 1998) e a imagem binária criada a partir do mapa de saliência com $t = 0,25$.

Na imagem binária a região branca representa o *foreground* e a escura representa o *background* da imagem. Existem diversas maneiras de definir o *threshold* t. No trabalho de (NAKAMOTO; TORIU, 2011), são utilizados alguns *thresholding* percentuais fixos (10%, 25%, 50%, 75% e 100%) para efetuar a separação do *foreground* e *background*. Também pode-se utilizar a Eq. (19) que representa a média simples dos valores do MS para definir o limiar, como apresentado em (SOARES; SILVA; GULIATO, 2012).

$$t = \frac{1}{mn}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} MS(i,j)\right) \tag{19}$$

Onde o $MS_{m \times n}$ é o mapa de saliência de uma imagem I e os elementos da matriz $MS(i, j) \in \mathbb{R}$.

Uma vez definido o processo binário de separar o *foreground* do *background* da imagem, o trabalho de (SOARES; SILVA; GULIATO, 2012), (baseado no (NAKAMOTO; TORIU, 2011)) propôs o descritor baseado na extração de característica utilizando o mapa de saliência binária. O mapa de saliência binária trabalha em conjunto com a abordagem BoVW, combinado com o descritor SIFT para representar os pontos-chave. Para construir o dicionário de palavras visuais, foi utilizado o algoritmo *k-means* (LIBERTY; SRIHARSHA; SVIRIDENKO, 2014; ELKAN, 2003; FORGY, 1965) e o mapa de saliência binária é usado para representar a localização espacial de palavras visuais na imagem.

A abordagem BoVW utilizando o mapa de saliência binária, dá origem à técnica denominada de *Binary Image Descriptor Saliency Map* (BSM). O BSM funciona da seguinte forma: primeiramente, cria-se o dicionário de palavras visuais D. Em seguida é extraído o MS (utilizando o modelo de (ITTI; KOCH; NIEBUR, 1998)) da imagem I e aplica-se a binarização no MS, escolhendo o t de acordo com a Eq. (19) ou com a abordagem (NAKAMOTO; TORIU, 2011). Após aplicar o limiar, será gerada uma imagem binária I_b que distingue o *foreground* e o *background*. Depois, aplica-se o modelo de (LOWE, 2004) para extrair os SIFTs da imagem I, e verifica-se se os mesmos estão localizados na parte do *foreground* ou *background* da imagem de acordo com I_b, ou seja, todos SIFTs de I que possuir o valor de seu respectivo MS maior que o *threshold* pertencerá ao *foreground* caso contrário ao *background*. Para criar os histogramas, são mapeados os SIFTs para as palavras visuais mais próximas, assim incrementada a sua frequência nos histogramas de palavras visuais do *foreground* ou do *background* de I. No estudo comparativo proposto neste trabalho, foi considerado apenas o *foreground* para dar maior importância às palavras que mais discriminam o objeto em análise. A Figura 16 ilustra o processo descrito anteriormente, considerando apenas o *foreground*.

A seguir, é apresentada uma outra abordagem que classifica o *pixel* localizado em regiões de transição do MS como *foreground* e *background* ao mesmo tempo. Esta abordagem é chamada de *Fuzzy Descriptor Image Saliency Map* (FISM).

3.2.3 *Fuzzy Descriptor Image Saliency Map* (FISM)

O método BSM apresentado demonstra fatores limitantes para encontrar um determinado *threshold* que otimize a classificação dos SIFTs em *foreground* e *background*. Uma alternativa é adotar um *threshold* que apresente o melhor desempenho. Mas é interessante considerar casos em que um determinado *pixel* em um MS pode estar localizado na transição entre a região *foreground* e *background*, fazendo com que o *pixel* se torne difícil de classificar. Um exemplo está ilustrado na Figura 17, que apresenta uma imagem e seu respectivo MS, o qual contém três objetos que destacam as regiões de *foreground*, *background* e a transição do *background* para o *foreground*.

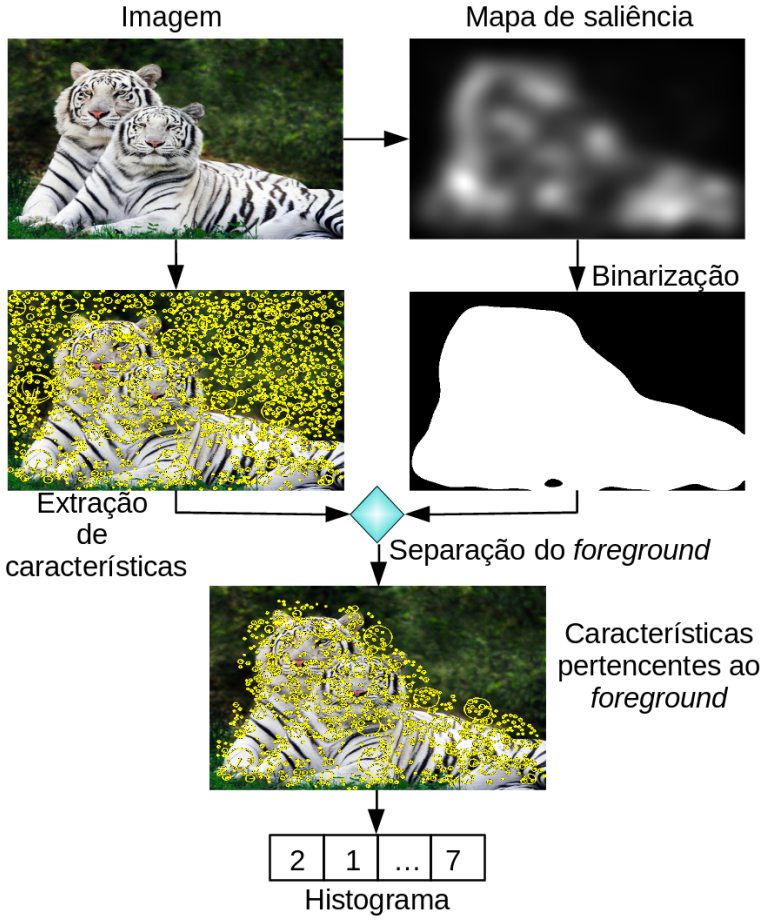

Figura 16 – Esquema para a geração dos descritores utilizando *Binary Saliency Map*.

Modificado de: (SOARES; SILVA; GULIATO, 2012).

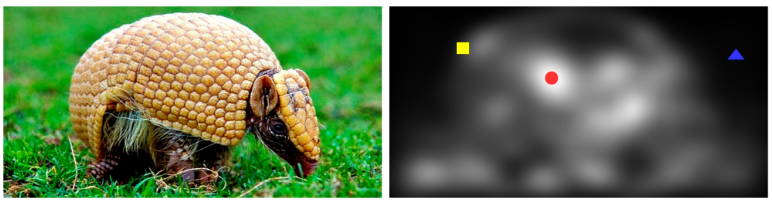

Figura 17 – Uma imagem e seu respectivo mapa de saliência contendo três objetos em destaque, o triângulo azul que representa a região de *background*, o círculo vermelho que expressa a região de *foreground* e o quadrado amarelo que está localizado na região de transição entre o *foreground* e o *background*.

Modificado de: (SOARES; SILVA; GULIATO, 2012).

Com base na análise dos *pixels* estarem localizados em regiões de transições, (SOARES; SILVA; GULIATO, 2012) propuseram um método para classificar um *pixel* como *foreground* e *background* ao mesmo tempo, permitindo modelar o grau de incerteza utilizando a Teoria dos Conjuntos Nebulosos (TCN). Nessa teoria, um elemento pertence a mais de um conjunto com distintos graus de pertinência. Deste modo, pode-se representar o grau de pertinência do *pixel* em uma determinada região no MS para *foreground* e para

o mesmo *pixel* um grau de pertinência para o *background*. Uma vez obtido o MS de uma imagem I, pode-se normalizar os valores do $MS(i,j)$ para o intervalo de $[0,1]$. O MS normalizado representa o grau de pertinência que cada *pixel* da imagem pertence a região *foreground*, enquanto que ao gerar o complemento deste MS, consegue-se obter o grau de pertinência de cada *pixel* que pertence ao *background* da imagem. A Eq. (20) mostra como achar o complemento de um único *pixel*. Repetindo esta operação para todos os *pixels* obtém-se a imagem complementar do MS. Deste modo, um *pixel* que tem o grau de pertinência de 70% como sendo do *foreground*, teria também o grau de pertinência de 30% como sendo da região *background*.

$$\overline{MS}(i,j) = 1 - MS(i,j) \tag{20}$$

Utilizando a TCN, não é mais necessário definir um *threshold* para separar as regiões do MS em *foreground* e *background*. E assim surge uma nova forma de criar um descritor utilizando a TCN que será denominada de FISM (SOARES; SILVA; GULIATO, 2012). Assim como BSM, o FISM também utiliza o descritor SIFT para representar os pontos-chave e o *k-means* para a construção do dicionário de palavras visuais. A diferença entre o BSM e o FISM são o modo de como será representado o *background* e o *foreground* de uma imagem. O FISM utiliza o processo de distinção *fuzzy* para gerar dois histogramas de frequências de palavras visuais, um histograma para os SIFTs que estão na região do *foreground* e outro histograma para os SIFTs que aparecem na região do *background* com um determinado grau de pertinência.

A abordagem geral do FISM é descrita a seguir. Primeiramente, cria-se o dicionário de palavras visuais D utilizando as imagens que estão no banco. Em seguida, aplica-se o descritor SIFT proposto por (LOWE, 2004) na imagem I para extrair suas características. Em seguida atribua-se a cada descritor SIFT a palavra visual mais próxima em D. Cria-se o MS utilizando o modelo (ITTI; KOCH; NIEBUR, 1998) como também o seu complemento $\overline{MS}$ da imagem I. Nessa etapa, será montado o histograma de *foreground* H_f que representará a frequência de características que aparecem no *foreground* da imagem I a partir da função de pertinência do MS. O H_f conterá um ponderamento da ocorrência do SIFT de acordo com a função de pertinência do MS. O mesmo procedimento descrito para o H_f será aplicado para o histograma do *background* H_b. Em vez de utilizar o MS para realizar o ponderamento, será utilizado o seu complemento $\overline{MS}$. Por fim, concatenam-se os dois vetores H_f com o H_b para criar o novo descritor. A Figura 18 apresenta um exemplo do esquema do descritor FISM.

O FISM descreve o *foreground* e o *background* das imagens separadamente. Esta separação permite aplicar ponderamentos nos histogramas H_f e H_b com o intuito de enfatizar a parte da consulta da imagem que é mais interessante a ser trabalhada para pesquisa de similaridade. Também é possível desconsiderar o H_f ou H_b ou simplesmente realizar a união de ambos, possibilitando, deste modo, adequar o FISM de acordo com as

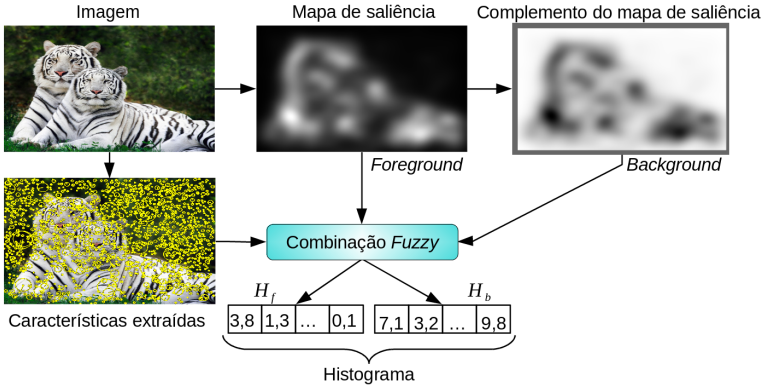

Figura 18 – Modelo para geração dos descritores FISM utilizando o mapa de saliência proposto por (ITTI; KOCH; NIEBUR, 1998).

Modificado de: (SOARES; SILVA; GULIATO, 2012).

características do banco que está sendo trabalhado.

As abordagens FISM e BSM são uma combinação do BoVW com o uso da técnica de atenção visual para simular computacionalmente o que é relevante ou não na imagem de acordo com a percepção humana. A seguir, é apresentada uma outra abordagem que utiliza o método BoVW juntamente com o Casamento por Pirâmides Espaciais (*Spatial Pyramids Matching* – SPM) para guardar informação espacial referente à imagem, denominada de *Bag-of-Visual-Words* com *Spatial Pyramids Matching* (BoVW-SPM).

3.3 BoVW com o Casamento por Pirâmides Espaciais (BoVW-SPM)

No modelo clássico do BoVW, a informação espacial referente a imagem não é armazenada e, para contornar esse problema, a abordagem BoVW foi combinada com o método SPM para auxílio na construção do histograma, denominada de *Bag-of-Visual-Words* com *Spatial Pyramids Matching* (BoVW-SPM).

A abordagem BoVW-SPM funciona da seguinte forma: primeiramente, utiliza um descritor (por exemplo o SIFT) para extração das características de cada região da imagem e posteriormente, na etapa de construção do dicionário visual usa-se um método de agrupamento (como exemplo, *k-means*). Em seguida, aplica-se o método SPM (LAZEBNIK; SCHMID; PONCE, 2006) para geração do histograma. O histograma criado pelo método SPM corresponde à união de vários outros histogramas gerados a partir de diferentes sub-regiões da imagem. E, estas sub-regiões podem ser divididas novamente criando novos histogramas. As sub-regiões são criadas com um critério de níveis e, quanto maior o nível, mais a imagem é segmentada em regiões. Ao final, com a concatenação de todos os histogramas que representam as diferentes regiões em diferentes níveis, forma-se uma representação única para a imagem.

O método *Spatial Pyramids Matching* – SPM), ou Casamento por Pirâmides Espaciais foi criado por Lazebnik (LAZEBNIK; SCHMID; PONCE, 2006) e baseado no método *pyramid match kernel* de Grauman (GRAUMAN; DARRELL, 2005). A técnica SPM na área de visão computacional tem sido amplamente utilizada para incorporar as informações espaciais globais e locais de uma imagem dentro de um vetor de característica (KRISTO; CHUA, 2013; PENG et al., 2014). A alternativa de se aplicar SPM tem conseguido um ganho na acurácia da classificação em aplicações de reconhecimento de objetos (LAZEBNIK; SCHMID; PONCE, 2006).

Em particular, o funcionamento da técnica SPM pode ser da seguinte maneira: a imagem é divida em uma sequência de grades cada vez mais finas em cada nível da pirâmide. Em um nível inicial (nível 0) a imagem original permanece sem divisões (contendo apenas uma região[2]). No nível seguinte (nível 1), subdivide a única região do nível 0 em 4 outras regiões de tamanhos similares (quadrantes), obtendo 4 histogramas. No nível 2, subdivide cada uma das regiões do nível 1 em 4 outras regiões, tendo assim um total de 16 regiões neste nível e, consequentemente, 16 histogramas que representam cada região, e este processo se repete assim por diante.

Os histogramas de descritores são extraídos para todas as regiões das grades e ponderados de acordo com as correspondências que ocorrem em cada nível. Em qualquer nível, dois pontos são ditos correspondentes se eles ocorrerem no mesmo *bin* da grade. É dado maior peso para casamentos que ocorrem nos níveis mais altos, ou seja, em regiões menores, refletindo assim o fato de que maiores níveis localizam as características mais precisamente. Por fim, são concatenados os histogramas de diferentes níveis para formar um único vetor que representará a imagem. A Figura 19 mostra um exemplo do uso da pirâmide espacial em uma imagem.

Formalizando esta abordagem, ao construir uma sequência de grades nas resoluções $0, ..., L$, o nível l da grade tem 2^l regiões em cada dimensão, para um total de $D = 2^{2l}$ regiões em cada nível. Seja I_1 e I_2 duas imagens, o H_I^l é o histograma de descritores da imagem I no nível l. O casamento para o nível l entre as imagens I_1 e I_2 pode ser calculado

$$C^l(I_{I_1}^l, I_{I_2}^l) = \sum_{i=1}^{2^{2l}} \min(H_{I_1}^l(i), H_{I_2}^l(i)), \tag{21}$$

Os casamentos do nível l incluem os casamentos que acontecem no nível $l+1$. Deste modo, os casamentos no nível l são dados por $C^l(I_{I_1}^l, I_{I_2}^l) - C^{l+1}(I_{I_1}^{l+1}, I_{I_2}^{l+1})$, para $l = 0, ..., L-1$. Os pesos são inversamente proporcionais à largura das regiões em cada nível. No nível l associa-se o peso $\frac{1}{2^{L-1}}$, deste modo terá um peso maior aos casamentos que acontecem nos níveis mais altos (que tem as regiões menores). A função núcleo para o

[2] Alguns autores utilizam o termo célula para referenciar a divisão que ocorre na imagem.

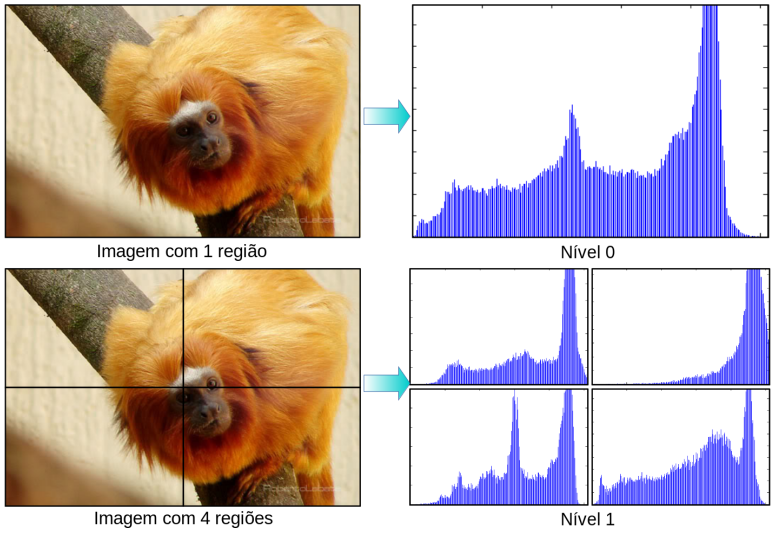

Figura 19 – Representação da Pirâmide Espacial com as divisões das regiões e seus respectivos histogramas nos níveis 0 e 1.

Imagem modificada de: (LAZEBNIK; SCHMID; PONCE, 2006)

.

casamento de pirâmides é dada por:

$$p^L(I_1, I_2) = \frac{1}{2^L} C^0(I^0_{I_1}, I^0_{I_2}) + \sum_{l=1}^{L} \frac{1}{2^{L-l+1}} C^l(I^l_{I_1}, I^l_{I_2}). \tag{22}$$

Ao aplicar a Eq. (22) para cada uma das N palavras visuais do dicionário (seção 3.1), o mecanismo da SPM pode ser escrito desta forma:

$$P^L(I_1, I_2) = \sum_{j=1}^{N} p^L(I_{1_j}, I_{2_j}) \tag{23}$$

onde I_j representa as coordenadas das palavras visuais do dicionário encontradas nas respectivas imagens.

Apesar de sua simplicidade, a SPM agrega confiança à informação espacial global, permite uma melhora no método de representação de imagens BoVW.

Geralmente, utilizam-se as medidas Cosseno e Euclidiana para calcular a similaridade dos histogramas gerados pelas abordagens BoVW, BSM, FISM e BoVW-SPM. Neste trabalho foram utilizadas estas mesmas abordagens, substituindo as funções de similaridade pela divergência de Bregman. No próximo Capítulo são apresentadas de forma sucinta algumas divergências de Bregman, com suas propriedades e alguns trabalhos correlatos que utilizam as DB's no contexto de recuperação de imagem baseada em conteúdo.

Capítulo 4

Divergência de Bregman

As divergências de Bregman (DB) foram introduzidas em 1967 pelo matemático L.M. Bregman (BREGMAN, 1967). Bregman tinha como principal preocupação de encontrar um ponto comum de conjuntos convexos, a fim de solucionar um problema de otimização convexa.

Para uma leitura mais agradável, será abordado antecipadamente o significado sobre algumas notações nesta seção: as variáveis $\mathbf{x}$ e μ são utilizadas para representar vetores. Um conjunto é representado pelo alfabeto caligráfico de letras maiúsculas $\mathcal{X}$, $\mathcal{Y}$. As variáveis aleatórias são expressas pelo alfabeto de letras maiúsculas, por exemplo X e Y. Os símbolos $\mathbb{R}$, $\mathbb{N}$, $\mathbb{Z}$ e $\mathbb{R}^d$ denotam os conjuntos dos reais, naturais, inteiros e o espaço vetorial real de dimensão-d respectivamente. Além disso, $\mathbb{R}_+$ e $\mathbb{R}_{++}$ indicam o conjunto não negativo e números reais positivos. Para $\mathbf{x}$ e $\mathbf{y} \in \mathbb{R}^d$, $\|\mathbf{x}\|$ expressa a norma L_2 e $\langle \mathbf{x}, \mathbf{y} \rangle$ indica o produto interno. O *log* representará os logaritmos naturais. A função de densidade de probabilidade (com relação à Lebesgue ou a medida contagem) são denotados pelo alfabeto de letras minúsculas tais como p e q. O interior relativo[1] de um conjunto convexo de $\mathcal{X}$ é denotado por $\text{ri}(\mathcal{X})$. O domínio efetivo de uma função f (por exemplo, o conjunto para todo x tal que $f(x) < +\infty$) é representado por $\text{dom}(f)$. A função inversa de f é denotada por f^{-1}.

É definida a divergência de Bregman correspondendo a uma função estritamente convexa da seguinte maneira (BREGMAN, 1967; CENSOR; ZENIOS, 1997):

Definição 1 Seja $\phi : \mathcal{S} \to \mathbb{R}$, uma função estritamente convexa definida em um conjunto convexo $\mathcal{S} \subseteq \mathbb{R}^d$ onde $\mathcal{S} = \text{dom}(\phi)$ tal que, ϕ é diferenciável em $\text{ri}(\mathcal{S})$ e não vazio. A **divergência de Bregman** $d_\phi : \mathcal{S} \times ri(\mathcal{S}) \to [0, \infty)$ entre dois elementos $\mathbf{x}$ e $\mathbf{y}$ é dada por:

$$d_\phi(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) - \phi(\mathbf{y}) - \langle \mathbf{x} - \mathbf{y}, \nabla\phi(\mathbf{y}) \rangle \tag{24}$$

onde $\nabla\phi(\mathbf{y})$ representa o vetor gradiente de ϕ avaliado em $\mathbf{y}$.

[1] Para todos os conjuntos convexo não vazios $\mathcal{C} \subseteq \mathbb{R}^d$ o interior relativo pode ser definido como $\text{ri}(\mathcal{C}) := \{x \in \mathcal{C} : \forall y \in \mathcal{C}\ \exists \lambda > 1 : \lambda x + (1 - \lambda) y \in \mathcal{C}\}$ (ROCKAFELLAR, 1996).

O $d_\phi(\mathbf{.},\mathbf{y})$ pode ser visto como a diferença entre ϕ e a aproximação dada pela sua expansão em série de Taylor ao redor de $\mathbf{y}$. Na Figura 20, a interpretação geométrica da divergência de Bregman no espaço 1-dimensional $\mathcal{S}$ é ilustrado. Nesta figura, a curva da função $Y' = \phi(X')$ e a reta $h : Y' = \phi(\mathbf{y}) + \langle \nabla\phi(\mathbf{y}), X' - \mathbf{y}\rangle$ passam pelo ponto $(\mathbf{y}, \phi(\mathbf{y}))$, são plotados separadamente. Note que h é a reta tangente a curva ϕ no ponto vermelho. A diferença vertical na posição $X' = \mathbf{x}$ (respectivamente $X' = \mathbf{r}$) é a medida da divergência de Bregman, $d_\phi(\mathbf{x}, \mathbf{y})$ (respectivamente $d_\phi(\mathbf{r}, \mathbf{y})$).

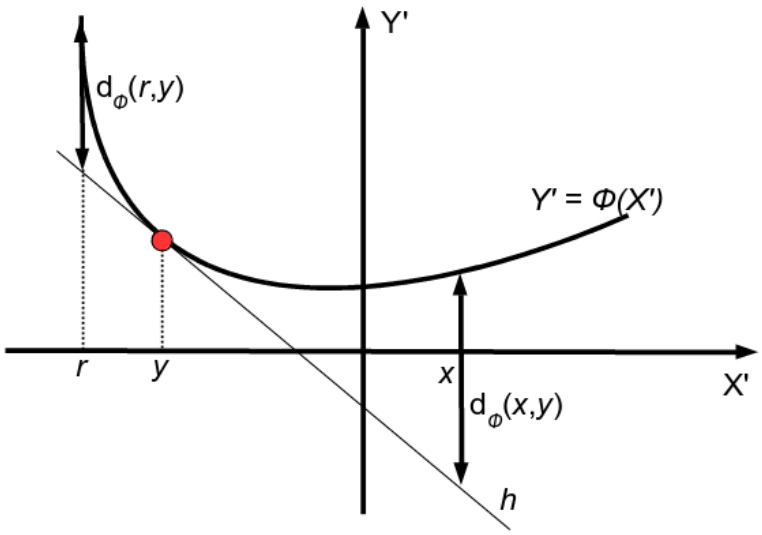

Figura 20 – Interpretação geométrica da divergência de Bregman.

Observe que diferentes escolhas para a função ϕ induzem diferentes métricas. Por exemplo, a distância Euclidiana, divergência de Kullback-Leibler (KL), a distância Mahalanobis e muitas outras distâncias amplamente utilizadas são casos especiais da divergência de Bregman, obtidas a partir da escolha de diferentes funções ϕ (BANERJEE et al., 2005). As seguintes propriedades são verdades de acordo com a Definição 1:

- **Não negatividade**: $d_\phi(\mathbf{x},\mathbf{y}) \geq 0$, $\forall \mathbf{x} \in \mathcal{S}$, $\mathbf{y} \in \text{ri}(\mathcal{S})$, e $d_\phi(\mathbf{x}, \mathbf{y}) = 0$ se e somente se $\mathbf{x} = \mathbf{y}$.

- **Convexidade**: d_ϕ é sempre convexo no primeiro argumento, mas não necessariamente convexo no segundo argumento. A distância *Square Euclidean* e *KL-Divergence* 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 $\phi(x) = x^3$, definida em $\mathbb{R}_+$, dado por $d_\phi(x, y) = x^3 - y^3 - 3(x - y)y^2$.

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

$$d_{\phi_1+\phi_2}(\mathbf{x}, \mathbf{y}) = d_{\phi_1}(\mathbf{x}, \mathbf{y}) + d_{\phi_2}(\mathbf{x}, \mathbf{y}) \ ,$$

$$d_{c\phi}(\mathbf{x},\, \mathbf{y}) = c d_{\phi}\ (\mathbf{x},\, \mathbf{y}),\ (\text{para } c \geq 0)\ .$$

- ❑ **Classes de equivalência**: As divergências de Bregman de funções que diferem apenas em termos afins são idênticos, ou seja, se $\phi(\mathbf{x}) = \phi_0(\mathbf{x}) + \langle \mathbf{b},\, \mathbf{y} \rangle + c$ onde $\mathbf{b} \in \mathbb{R}^d$ e $c \in \mathbb{R}$, depois $d_{\phi}(\mathbf{x},\mathbf{y}) = d_{\phi_0}(\mathbf{x},\mathbf{y})$, $\forall \mathbf{x} \in \mathcal{S}$, $\mathbf{y} \in \text{ri}(\mathcal{S})$. Assim, o conjunto de todas as funções diferenciáveis estritamente convexas em um conjunto convexo $\mathcal{S}$, pode ser dividida em classes de equivalência na forma:

$$[\phi_0] = \{\phi \mid d_{\phi}(\mathbf{x},\, \mathbf{y}) = d_{\phi_0}(\mathbf{x},\, \mathbf{y})\ \forall\ \mathbf{x} \in \mathcal{S},\, \mathbf{y} \in \text{ri}(\mathcal{S})\}.$$

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

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

Geralmente, as divergências de Bregman não são simétricas, por exemplo, a distância Euclidiana ao quadrado. A Tabela 2 contém uma lista de algumas divergências de Bregman e suas funções básicas. Vale destacar que na distância Mahalanobis a matriz A é assumida como sendo positiva definida; $(\mathbf{x} - \mathbf{y})^T A(\mathbf{x} - \mathbf{y})$ é chamado de distância Mahalanobis quando A é inversa da matriz de covariância.

Tabela 2 – Algumas funções convexas das divergências de Bregman.

Domínios	$\phi(\mathbf{x})$	$d_{\phi}(\mathbf{x}, \mathbf{y})$	**Divergência**
$\mathbb{R}$	x^2	$(x - y)^2$	*Squared loss*
$\mathbb{R}_{++}$	$-\log x$	$\frac{x}{y} - \log(\frac{x}{y}) - 1$	Distância Itakura-Saito
$\mathbb{R}^d$	$\|\mathbf{x}\|^2$	$\|\mathbf{x} - \mathbf{y}\ \|^2$	Euclidiana ao quadrado
$\mathbb{R}^d$	$\mathbf{x}^T A\mathbf{x}$	$(\mathbf{x} - \mathbf{y})^T A(\mathbf{x} - \mathbf{y})$	Distância Mahalanobis
d-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*

A seguir são apresentados alguns exemplos das divergências de Bregman (BANERJEE et al., 2005).

Exemplo 1. A distância *Squared Euclidean* é talvez a mais simples e mais amplamente usada das divergências de Bregman. A função base $\phi(\mathbf{x}) = \langle \mathbf{x},\mathbf{x} \rangle$ é estritamente convexo e diferenciável no $\mathbb{R}^d$ e

$$\begin{aligned}
d_{\phi}(\mathbf{x},\, \mathbf{y}) &= \langle \mathbf{x},\mathbf{x} \rangle - \langle \mathbf{y},\mathbf{y} \rangle - \langle \mathbf{x} - \mathbf{y},\, \nabla\phi(\mathbf{y}) \rangle \\
&= \langle \mathbf{x},\mathbf{x} \rangle - \langle \mathbf{y},\mathbf{y} \rangle - \langle \mathbf{x} - \mathbf{y},\, 2\mathbf{y} \rangle \\
&= \langle \mathbf{x} - \mathbf{y},\, \mathbf{x} - \mathbf{y} \rangle = \|\mathbf{x} - \mathbf{y}\|^2.
\end{aligned}$$

Exemplo 2. Outra divergência de Bregman usada é a KL. Se $\mathbf{p}$ é uma distribuição de probabilidade discreta de modo que $\sum_{j=1}^{d} p_j = 1$, a entropia negativa $\phi(\mathbf{p}) = \sum_{j=1}^{d} p_j \log_2 p_j$ é uma função convexa. A divergência de Bregman correspondente é:

$$\begin{aligned} d_\phi(\mathbf{p},\mathbf{q}) &= \sum_{j=1}^{d} p_j \log_2 p_j - \sum_{j=1}^{d} q_j \log_2 q_j - \langle \mathbf{p}-\mathbf{q}, \nabla\phi(\mathbf{q})\rangle \\ &= \sum_{j=1}^{d} p_j \log_2 p_j - \sum_{j=1}^{d} q_j \log_2 q_j - \sum_{j=1}^{d} (p_j - q_j)(\log_2 q_j + \log_2 e) \\ &= \sum_{j=1}^{d} p_j \log_2 \left(\tfrac{p_j}{q_j}\right) - \log_2 e \sum_{j=1}^{d} (p_j - q_j) \\ &= KL(\mathbf{p}\|\mathbf{q}), \end{aligned}$$

A divergência KL entre as duas distribuições como $\sum_{j=1}^{d} q_j = \sum_{j=1}^{d} p_j = 1$.

Exemplo 3. A distância *Itakura-Saito* é outra divergência de Bregman que é bastante utilizada em processamento de sinal. Se $F(e^{j\theta})$ é um espectro de energia[2] de um sinal $f(t)$, então o funcional $\phi(F) = -\frac{1}{2\pi}\int_{-\pi}^{\pi} \log(F(e^{j\theta}))d\theta$ é convexo em F e corresponde à taxa de entropia negativa de um sinal, assumindo que seria gerado por um processo Gaussiano fixo (PALUS, 1997).

A divergência de Bregman entre $F(e^{\theta})$ e $G(e^{\theta})$ (o espectro de energia de um outro sinal $g(t)$) é dado por:

$$\begin{aligned} d_\phi(F,G) &= \tfrac{1}{2\pi}\int_{-\pi}^{\pi}(-\log(F(e^{j\theta})) + \log(G(e^{j\theta})) - (F(e^{j\theta}) - G(e^{j\theta}))(-\tfrac{1}{G(e^{j\theta})}))d\theta \\ &= \tfrac{1}{2\pi}\int_{-\pi}^{\pi}(-\log(\tfrac{F(e^{j\theta})}{G(e^{j\theta})}) + \tfrac{F(e^{j\theta})}{G(e^{j\theta})} - 1)d\theta \qquad , \end{aligned}$$

que é exatamente a distância de *Itakura-Saito* entre a energia espectral $F(e^{j\theta})$ e $G(e^{j\theta})$ e pode também ser interpretada como a *I-divergence* (CSISZÁR, 1991), entre os processos de geração sob o pressuposto que a média é igual aos processos Gaussianos fixos (KAZAKOS, 2006).

Para a próximo seção, abordaremos alguns trabalhos da literatura que utilizam a divergência de Bregman como função de similaridade no contexto de recuperação de imagens.

4.1 Trabalhos correlatos: as divergências utilizadas no contexto de recuperação de imagens

Nesta seção estão descritos alguns trabalhos encontrados na literatura que utilizam as divergências de Bregman como medida de similaridade no contexto de recuperação de imagens baseada em conteúdo.

Destacamos, em primeiro lugar, o trabalho de (XU et al., 2012), que aplica a divergência CBIR. Na obra de (LIU et al., 2012), para recuperação de formas de objetos foi utilizada como medida de similaridade a divergência total de Bregman. Enquanto que

[2] Note que $F(.)$ é uma função e é possível estender a notação para divergência de Bregman para o espaço de funções (CSISZÁR, 1995) e (GRUNWALD; DAWID, 2004)

o artigo de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) descreve sobre a utilização da distância KL simétrica como medida de similaridade em sistema CBIR, em que no processamento *off-line* do banco de dados foi aplicada a *Wavelet Transform* (WT) para caracterizar as imagens do banco, enquanto que no modo *on-line* foi realizada a consulta utilizando a distância KL simétrica. No trabalho de (SCHWANDER; NIELSEN, 2010), também fundamentado no campo de recuperação de imagem baseada em conteúdo, os autores substituíram o uso da distância *squared Euclidean* utilizada como sub-rotina do *k-means* pelas α-divergentes. A pesquisa de (PIRO et al., 2008) aplicou a divergência KL como medida de similaridade entre distribuições marginais parametrizadas de coeficientes *wavelet* em diferentes escalas.

O trabalho de (KWITT; UHL, 2008) apresentou uma aplicação para recuperação de imagem baseada em textura, com intuito de utilizar a KL para medir as distribuições marginais das magnitudes de coeficientes complexo de *wavelet*. Por fim, no trabalho de (BANERJEE et al., 2005) foi proposta uma análise paramétrica *hard* e *soft* de algoritmos de agrupamentos baseados nas divergências de Bregman.

Para descrever melhor sobre os trabalhos correlatos existentes na literatura, detalhamos nas próximas subseções três propostas citadas anteriormente que utilizam as divergências no contexto de similaridade, que são: os trabalhos de (XU et al., 2012), (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) e (KWITT; UHL, 2008).

4.1.1 Trabalho 1 – proposto por (XU et al., 2012)

O trabalho de (XU et al., 2012) utiliza algoritmos de clusterização supervisionados e não-supervisionados baseados em ranqueamento de gráfico. Esses algoritmos de clusterização têm recebido considerável atenção em aprendizado de máquina, visão computacional e na comunidade de pesquisadores de recuperação de informação. A classificação no coletor de dados (ou classificação do coletor (*manifold ranking* (MR))) é uma das aproximações representativas e tem sido amplamente aplicado em várias aplicações de recuperações de informação e aprendizado de máquina. Na pesquisa desenvolvida por (XU et al., 2012) foi feito um modelo de MR dentro de um *framework* otimizado utilizando a divergência de Bregman, o qual transformou o MR em um problema de aprendizado da matriz do "*kernel*" ótimo. Com esta nova formulação, duas extensões eficientes e efetivas são propostas para aumentar a performance do ranqueamento.

A ideia central da classificação do coletor (MR) é ranquear o dado com relação à estrutura geométrica intrínseca revelada coletivamente por uma grande quantidade de dados (sem rótulo). Considerando ambas, tanto a rotulada (a consulta) e a não rotulada da base (o banco de dados), atribui a cada ponto de dados um valor de ranqueamento relativo, o qual pode ser considerado como o grau de relevância para a consulta. Caso contrário, a similaridade entre pares ou as distâncias usadas em muitos métodos tradicionais, o valor

do ranqueamento é mais significativo para medir a relevância semântica expressada dentro da estrutura geométrica subjacente do conjunto de dados.

Assim, um dos principais inconvenientes do MR é sua alta complexidade. Dada uma *query* tanto a construção do MR de um gráfico de afinidade e a propagação dos valores de ranqueamentos no gráfico leva uma complexidade de $O(n^3)$, onde n é o número de amostras no banco de dados. Se a consulta se encontra no banco de dados, o MR pode utilizar a pré computação *off-line* para reduzir o custo *on-line*. Porém, para uma consulta fora do banco de dados, o alto custo da etapa de propagação do valor de ranqueamento precisará ser realizada no estágio *on-line*, o qual é usualmente referido como o problema fora da amostra (*out-of-sample*).

As principais contribuições do trabalho de (XU et al., 2012) foram: (1) a formulação de um algoritmo de ranqueamento de "*manifolds*" para um problema de otimização utilizando divergência de Bregman; (2) possibilidade de um novo entendimento no contexto de MRs (*manifold ranking*) introduzindo "aprendizado" usando uma matriz de *kernel* e utilizando a divergência de Bregman; e (3) com a nova formulação, são propostas duas extensões eficientes e efetivas para melhorar a performance da tradicional MR denominadas de *efficient divergence view of manifold ranking* (DMR_E) e *Constraints divergence view of manifold ranking* (DMR_C).

O DMR_C permite utilizar a informações de restrições de pares induzidos dos *feedbacks* do usuário para guiar o ranqueamento, sendo uma maneira promissora para algoritmos de ranqueamento semi-supervisionado.

O *Manifold Ranking* (MR), pode ser definido da seguinte forma: dado um conjunto de dados $X = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ onde cada coluna é uma amostra de vetor de tamanho m. O MR primeiro constrói um gráfico de afinidade sobre os dados (por exemplo, gráfico *k-Nearest Neighbors* (k-NN)). Seja $W \in \mathbb{R}^{n \times n}$ que denota o peso da matriz do gráfico com w_{ij} armazenando o peso de cada aresta. Uma maneira comum para computar o peso é usando o *heat kernel* $w_{ij} = exp[-d^2(\mathbf{x}_i, \mathbf{x}_j)/2\sigma^2]$ se tem uma aresta ligando $\mathbf{x}_i$ e $\mathbf{x}_j$, caso contrário $w_{ij} = 0$. A função d é uma distância métrica, tal como a distância Euclidiana.

Seja $\mathbf{f}$ uma função de ranqueamento que atribui para cada ponto de $\mathbf{x}_i$ um valor de ranqueamento f_i. O MR define um vetor inicial $\mathbf{f}^0 = [f_1^0, ..., f_n^0]^T$, no qual $f_i^0 = 1$ se x_i é uma consulta e $f_i^0 = 0$ caso contrário. O custo da função associado com f em MR é definida em (ZHOU et al., 2004), por:

$$O(\mathbf{f}) = \frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij}\left(\frac{f_i}{\sqrt{D_{ii}}} - \frac{f_j}{\sqrt{D_{jj}}}\right)^2 + \mu\|\mathbf{f} - \mathbf{f}^0\|^2\right), \tag{25}$$

onde $\mu > 0$ é o parâmetro de regularização e D é a matriz diagonal com $D_{ii} = \sum_j w_{ij}$.

O algoritmo pode também ser designado como uma forma interativa como a seguir

$$\mathbf{f}(t+1) = \varrho S\mathbf{f}(t) + (1-\varrho)\mathbf{f}^0, \tag{26}$$

onde $S = D^{-1/2}WD^{-1/2}$ e $\varrho = \mu/(1+\mu)$. Durante uma interação t, cada ponto recebe informação dos seus vizinhos e mantém sua atribuição inicial. Por fim, o algoritmo converge para

$$\mathbf{f}^* = (I_n - \varrho S)^{-1}\mathbf{f}^0 = K\mathbf{f}^0. \tag{27}$$

É importante notar que o parâmetro ϱ deverá está no intervalo [0,1)[3], caso contrário o algoritmo não converge. O K é a matriz de pesos que pode ser pré-computada de forma *off-line* caso a consulta não esteja no banco de dados, ou *on-line* caso contrário.

Desta forma, o trabalho de (XU et al., 2012) formulou o algoritmo MR a partir do *framework* otimizado da divergência de Bregman e, algumas extensões foram derivadas para superar as deficiências da abordagem MR tradicional. Assim, a nova formulação do algoritmo MR pode ser da seguinte maneira: seja $Y = [\mathbf{y}_1, ..., \mathbf{y}_n] \in \mathbb{R}^{p\times n}$ a representação dos dados em um novo espaço de características das amostras dos dados. É o mesmo que falar $\mathbf{y}_i = \Psi(\mathbf{x}_i)$, para $i = 1, ..., n$, onde Ψ é uma função de transformação do dado para um novo espaço de característica. A matriz K é definida como

$$K = Y^T Y. \tag{28}$$

A matriz K é sempre semi-definida desde que mantém para qualquer vetor $\mathbf{v}$, $\mathbf{v}^T K\mathbf{v} = \|Y\mathbf{v}\|^2 \geq 0$. Esta matriz K na formulação do MR (Eq. (27)) é a solução para o seguinte problema de otimização

$$\min_K D_{ld}(K, I) \tag{29}$$

de modo que

$$\sum_{i,j} \|\frac{1}{\sqrt{D_{ii}}}\mathbf{y}_i - \frac{1}{\sqrt{D}_{jj}}\mathbf{y}_j\|^2 w_{ij} \leq \delta, K \succeq 0, \tag{30}$$

onde δ é um parâmetro de controle de suavidade da restrição que faz as amostras vizinhas terem curtas distâncias no novo espaço. O $D_{ld}(K, I)$ é o resultado da matriz de divergência de Bregman, onde $D_{ld}(K, I) = tr(KI^{-1}) - \log\det(KI^{-1}) - n$, sendo denominada de divergência *Log-Determinant* (KULIS; SUSTIK; DHILLON, 2006). A demonstração está apresentado em (XU et al., 2012).

A otimização da divergência de Bregman apresentada em (XU et al., 2012) atendendo o MR mostra que é possível encontrar a matriz ótima K "*closest*" para identificar a matriz sob certas restrições. Foi nomeado esta formulação como *divergence view of MR* (DMR). Deste modo, é possível fazer uma nova extensão do MR que seja eficiente.

Para identificar a matriz K "*closest*" foi utilizado a matriz *Gaussian kernel*, cada elemento da matriz é calculado por

$$K_{ij}^G = exp(-d^2(\mathbf{x}_i, \mathbf{x}_j)/2\sigma^2), \tag{31}$$

[3] Onde $\varrho = 0$, $\mathbf{f}^*$ é sempre igual ao $\mathbf{f}^0$ inicial.

onde σ é o parâmetro do tamanho da janela e $d(\mathbf{x}_i, \mathbf{y}_i)$ retorna a distância Euclidiana entre $\mathbf{x}_i$ e $\mathbf{x}_j$.

O trabalho de (XU et al., 2012) conseguiu reduzir a complexidade de $O(n^3)$ para $O(n)$, sendo mais eficiente do que o tradicional algoritmo MR. Este algoritmo é denominado de *efficient* DMR (DMR_E). Também foi realizada uma segunda extensão, que utiliza o *pairwise constraints*, que tem sido amplamente estudado em clusterização semi-supervisionado em trabalhos de aprendizagens métricas denominado de DMR com *Constraints* (DMR_C).

Em (XU et al., 2012) foram feitos vários experimentos que utilizam o método DMR_E, comparados com outros métodos de classificação não supervisionados, utilizando o banco de dados Caltech101. Para caracterização foram extraídos quatro tipos de características efetivas, resultando para cada imagem um vetor de tamanho 297. O primeiro tipo de característica foi *Grid Color Moment*, que particionava a imagem em grade de 3×3. Para cada grade foi extraído a média, variância e a assimetria de cada canal de cor (R, G e B – *Red*, *Green* e *Blue*) respectivamente, gerando assim um vetor de tamanho 81 correspondendo ao momento de cor. A segunda foi o *edge detector* proposto por (CANNY, 1986) que resultou um vetor de tamanho 37 para cada imagem. O terceiro tipo de característica foi o *gabor wavelets texture*, que aplica a transformada de *wavelet Gabor* (LADES et al., 1993) em imagens redimensionadas para 64×64, aplicado com 5 níveis e 8 orientações, que resultou em 40 sub-imagens. Para cada sub-imagem 3 momentos são calculados, média, variância e assimetria, ao final resultou em um vetor de tamanho 120. A última caracterização foi a *Local Binary Pattern* (OJALA; PIETIKÄINEN; HARWOOD, 1996), que consiste na medida de níveis de cinza, derivando da definição da textura geral em uma vizinhança local, gerando assim um vetor de tamanho 59.

Para comparação dos resultados, os autores utilizaram a Eud (distância Euclidiana) como *baseline*, o *Principal Component Analysis* (PCA) para redução de dados, o Mah que é o método padrão para a distância métrica que utiliza uma amostra da matriz de covariância, MR e o DMR_E. As métricas de avaliação do sistema CBIR foram o nDCG, MAP das 200 primeiras imagens (MAP@200) retornadas. A performance do sistema é a média sobre todas as consultas. O experimento dos autores foi conduzido no banco de dados Caltech101 (veja a seção 7.1) o qual, utilizou todas as imagens do banco como consulta e os resultados obtidos estão apresentados na Tabela 3.

Analisando os resultados obtidos por (XU et al., 2012), verifica-se que a recuperação utilizando a Eud e o MR não retornaram bons resultados e, o DMR_E supera significantemente todos os outros métodos não supervisionados.

Apresentamos, em seguida, o trabalho proposto pelos autores (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010).

Tabela 3 – Comparação da performance **sem** restrição no banco Caltech101. Para cada resultado, registrou-se o valor médio (%) de todas as consultas e a melhoria relativa sobre o método *baseline* Eud. O melhor resultado em cada linha é indicado pela fonte em negrito.

Tabela modificada de: (XU et al., 2012).

	Métodos sem restrições				
Método	Eud	PCA	Mah	MR	DMR_E
MP@10	35,50	34,03 -4,15%	37,34 +5,18%	34,98 -1,48%	**38,93 +09,65%**
MP@20	32,52	31,27 -3,86%	33,75 +3,77%	32,47 -0,15%	**35,72 +09,85%**
MP@30	30,73	29,55 -3,83%	31,54 +2,63%	30,88 +0,47%	**33,73 +09,77%**
nDCG@10	37,28	35,63 -4,43%	39,26 +5,30%	36,79 -1,31%	**40,77 +09,73%**
nDCG@20	34,55	33,12 -4,15%	36,06 +4,35%	34,38 -0,50%	**37,87 +09,60%**
nDCG@30	32,85	31,51 -4,08%	33,99 +3,50%	32,84 -0,02%	**36,00 +09,59%**
MAP@200	33,25	31,56 -5,07%	34,46 +3,66%	35,23 +5,98%	**36,65 +10,25%**

4.1.2 Trabalho 2 – proposto por (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010)

Neste trabalho, os autores utilizam a divergência KL Simétrica (Kullback Leibler Simétrica (KLS)) como medida de similaridade no contexto de recuperação de imagem baseada em conteúdo. O sistema CBIR proposto foi constituído em duas partes. A primeira parte é o processamento *off-line*, o qual é aplicado a *Wavelet Transform* (WT) nas imagens do banco de dados, em que as assinaturas relevantes são computadas a partir dos resultados dos coeficientes *wavelet*. A segunda parte é composta de um procedimento *on-line* que consiste na consulta e na recuperação das imagens cujas assinaturas são mais semelhantes, usando uma determinada medida de similaridade. No trabalho destes autores, a distância KLS, foi escolhida para avaliar o grau de semelhança entre a distribuição da imagem de consulta com alguma imagem da base.

Na etapa de extração de características foram utilizados os *Lifting Schemes* (LS), que são ferramentas convenientes para construir qualquer *wavelet* bi-ortogonal compactamente suportada. No caso 1D, o primeiro passo do LS é dividir o sinal $a_0(n)$ com $n \in \mathbb{Z}$ em suas amostras de pares e ímpares. Em seguida, as amostras pares $a_0(2n)$ são previstas a partir dos ímpares, o coeficiente de predição residual ou WT coeficientes $d_1(n)$ são computados como $d_1(n) = a_0(2n) - \lfloor \mathbf{a}_0(n)^T \mathbf{p} \rceil$, onde $\lfloor \cdot \rceil$ é o operador de arredondamento, $\mathbf{a}_0(n)$ apenas contém algumas amostras ímpares de $a_0(n)$ e $\mathbf{p}$ é o vetor de pesos de predição. Finalmente, o sinal de aproximação $a_1(n)$ é obtido por atualizar as amostras ímpares por alguns coeficientes de detalhes $d_1(n) : a_1(n) = a_0(2n-1) + \lfloor \mathbf{d}_1(n)^T \mathbf{u} \rceil$ onde $\mathbf{d}_1(n)$ contém amostras dos coeficientes de detalhes $d_1(n)$ e $\mathbf{u}$ é o vetor de pesos atualizados. Esta decomposição é recursivamente aplicada para aproximação do sinal.

Para uma imagem de B-elementos, muito frequentemente, a decomposição é aplicada separadamente para cada elemento $b \in \{1, ..., B\}$ de tamanho $N \times N$. Para cada elemento

b, este procedimento se repete em J estágios, gera uma aproximação da imagem com uma resolução mais grosseira e $3J$ sub-bandas de *wavelet* de tamanho $N_j \times N_j$ para $j \in \{1, ..., J\}$, com $N_j = \frac{N}{2^j}$ horizontalmente, verticalmente e diagonalmente orientada. Por exemplo, a $\frac{5}{3}$ *transform* manteve o novo padrão de imagem sem perda da codificação padrão JPEG 2000 definida por:

$$\mathbf{p} = \left(\frac{1}{2}, \frac{1}{2}\right)^T, \quad \mathbf{a}_0(n) = (a_0(2n-1), a_0(2n+1))^T \tag{32}$$

$$\mathbf{u} = \left(\frac{1}{4}, \frac{1}{4}\right)^T, \quad \mathbf{d}_1(n) = (d_1(n-1), d_1(n))^T. \tag{33}$$

No trabalho de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) agrupou em três sub-bandas os detalhes em cada nível j de modo a obter uma única sub-banda w_j^b de tamanho $N_j \times 3N_j$.

O que motivou o trabalho de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) foi a existência de dependências espectrais significativas entre elementos e, por tanto, entre os seus coeficientes WT relacionados para definir o vetor B-variável $w_j(n,m) = (w_j^{(1)}(n,m), ..., w_j^B(n,m))^T$ para cada posição espacial (n,m) no interior da j-ésima sub-banda. Por esse motivo, o objetivo do trabalho de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) foi extrair algumas assinaturas salientes de um modelo paramétrico multivariado adequado P_j da distribuição conjunta de $\mathbf{w}_j$, assim como parâmetros subjacentes são considerados como assinaturas salientes da sub-banda B na escala j de uma dada imagem multi-espectral.

Para um procedimento de recuperação escalável foi necessário definir uma medida de similaridade entre a distribuição $\mathbf{P}_j^{i_{db}}$ de uma imagem i_{db} no banco de dados e a distribuição $\mathbf{P}_j^{i_q}$ da imagem de consulta na sub-banda j. Foi utilizada a divergência KL Simétrica (KLS), que é conhecida por ser uma distância D apropriada para comparar duas distribuições $\mathbf{P}_j^{i_q}$ e $\mathbf{P}_j^{i_{db}}$ (definida na seção **??**).

O procedimento de recuperação para uma imagem de consulta i_q consiste no resultado da imagem i_{db}^* que minimiza a KLS global:

$$i_{db}^* = \arg\min_{i_{db}} \sum_{j=J_u}^{J} D(\mathbf{P}_j^{i_{db}}, \mathbf{P}_j^{i_q}). \tag{34}$$

onde J_u é uma escala escolhida por utilizar $(1 \leq J_u \leq J)$. Mais precisamente, a partir do modelo de densidade $\mathbf{P}_j^{i_{db}}$, uma amostra aleatória $s_1, ..., s_{L_j^{i_{db}}}$ de tamanho $L_j^{i_{db}}$ é gerado e a estimativa $\hat{d}$ é computada como:

$$\hat{d}(\mathbf{P}_j^{i_{db}} \parallel \mathbf{P}_j^{i_q}) = \frac{1}{L_j^{i_{db}}} \sum_{l=1}^{L_j^{i_{db}}} \log(\mathbf{P}_j^{i_{db}}(\mathbf{s}_l)) - \log(\mathbf{P}_j^{i_q}(\mathbf{s}_l)). \tag{35}$$

Em seguida, a estimativa empírica $\hat{D}$ da divergência KLS D (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010):

$$\hat{D}(\mathbf{P}_j^{i_{db}}, \mathbf{P}_j^{i_q}) = \frac{1}{2}\left(\hat{d}(\mathbf{P}_j^{i_{db}} \parallel \mathbf{P}_j^{i_q}) + \hat{d}(\mathbf{P}_j^{i_q} \parallel \mathbf{P}_j^{i_{db}})\right) \tag{36}$$

Com isso, a proposta de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) consistiu em desenvolver uma organização eficiente (*off-line*) dos índices do banco de dados que permite uma busca escalável, assegurando uma boa precisão, como apresentado a seguir:

- ❏ No nível inicial j é definido como $j = J$ e, o número de classes K_J é inicializado e um vetor $\mathbf{k}_{J-1}$ de tamanho K_J é escolhido.
- ❏ No nível j, a distribuição $\mathbf{P}_j^{i_{db}}$ são classificadas em K_j classes aplicando a regra do vizinho mais próximo com base na distância mantida. O centróide de cada resultado da classe será considerada como o protótipo da classe. Se a *Normalized Euclidean Distance* (NED) é considerada, o protótipo reduz a média das versões normalizadas dos vetores de características $\mathbf{f}_j^{i_{db}}$.
- ❏ Se $j > 1$, então o próximo nível $(j + 1)$, o valor de j é decrementado 1. Para cada classe k de K_{j+1} classes no nível anterior, o procedimento de agrupamento só diz respeito as imagens dentro da classe k: suas distribuições ou vetores de características no nível atual j são divididos em subconjuntos seguindo o mesmo algoritmo de agrupamento. Portanto, o número inteiro de classes K_j é dado por: $K_j = \sum_{k=1}^{K_{j+1}} \mathbf{k}_j[k]$ onde $\mathbf{k}_j[k]$ denota o k-ésimo elemento de $\mathbf{k}_j$ e seu valor é o número de subclasses de classes k na escala mais grosseira $j + 1$.
- ❏ Se $j = 1$, o procedimento de agrupamento é parado.

Este método pode ser usado para um algoritmo de agrupamento não supervisionado. A forma de consulta utilizada foi o procedimento de recuperação *coarse-to-fine*. Ele inicia no mais grosseiro (*coarsest*) J e é interrompido (parado) na escala j_u escolhido pelo usuário. Mais precisamente, $\mathbf{f}_J^{i_q}$ é comparada com os K_J protótipos de acordo com a divergência KLS. É atribuído ao *cluster* mais próximo $k_J^{i_q}$. Então, $\mathbf{f}_{J-1}^{i_q}$ é comparado com os protótipos do subconjuntos relacionados a $k_J^{i_q}$ e assim por diante. Se $j = J_u$, a saída do sistema são R imagens mais próximas dentro do grupo atribuído.

Os experimentos conduzidos foram utilizados a base de treinamento de 1.000 imagens SPOT3, representando diferentes regiões da Tunísia. O WT foi aplicado para cada imagem do banco de dados. Como imagem de consulta foram utilizadas 300 imagens do banco de treinamento.

Muitas melhorias deveriam ser esperadas, considerando todos os componentes espectrais disponíveis. Um segundo estágio 5/3 WT foi aplicado para cada imagem no banco de dados ($J = 2$) e definiu-se $J_u = 1$, de modo que os dois estágios são usados durante

a recuperação. Como imagens de consulta de teste i_q, foram utilizados 300 imagens do banco de treinamento. Para cada imagem de consulta, o conjunto de imagens *ground truth* é definido como as imagens do banco dentro da mesma categoria que a imagem de consulta. A performance da recuperação são avaliadas em termos de precisão média $PR = R_r/R$ e *recall* médio $RC = R_r/R_t$, onde R_r é o número de saída de imagens consideradas como relevantes, R_t o número total de saídas das imagens consideradas como relevantes no banco de dados e R o número de imagens retornadas.

Assim, os experimentos do trabalho de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) foram separados em duas partes. A primeira parte dos experimentos apresenta o procedimento de recuperação de uma imagem de consulta i_q para CBIR e aplicar uma busca exaustiva, a fim de produzir as R imagens mais próximas i_{db} conforme a distância utilizada. Nesta etapa do experimento, estudou-se a influência da escolha do modelo de parâmetro da multivariada (densidades marginais e *copula density*) para o multi-componente dos coeficientes WT. A divergência KLS também foi utilizada para o cálculo de similaridade e garantiu um ganho significativo nos resultados.

A segunda parte dos experimentos visam estudar a performance de CBIR quando a proposta de busca escalável é realizada após a estruturação de árvore do banco de dados de índice com $K_2 = 3$, $\mathbf{k}_1[k] = 2$, para $k = 1, 2, 3$. Portanto, as características dos vetores na escala $j = J_u = 1$ são classificados em $K_1 = 6$ classes. Também são generalizados os experimentos para o caso em que $B = 3$, considerando todos os três elementos das imagens multi-espectral no banco de dados.

Os experimentos utilizando a árvore de busca estruturada e introduzindo a *density copula* melhorou a performance de CBIR. Porém, as performances da árvore de busca foram menos satisfeitas do que com a busca exaustiva. A busca exaustiva, ao contrário é baseada na comparação simétrica da consulta i_q com todas as imagens no banco de dados, e a busca escalável é limitada para a seleção das R imagens mais similares dentro do grupo atribuído $k_{J_u}^{i_q}$ no final da escala J_u. Entretanto, esta perda na performance foi compensada pelo ganho no número de computações (cálculos) da divergência KLS fornecido pela busca escalável.

O interessante do trabalho de (SAKJI-NSIBI; BENAZZA-BENYAHIA, 2010) é que o mesmo se baseia na organização inicial do banco de dados de índice, a fim de permitir uma busca rápida e escalonável na etapa de recuperação. Os experimentos provaram que a proposta da nova busca escalável reduz o tempo de recuperação e continua mantendo ao mesmo tempo desempenho aceitáveis durante a recuperação.

Por fim, a seguir, apresentamos o trabalho de (KWITT; UHL, 2008) que utiliza a divergência KL para medir a similaridade no contexto de recuperação de imagem baseado em textura.

4.1.3 Trabalho 3 – proposto por (KWITT; UHL, 2008)

O trabalho do (KWITT; UHL, 2008) apresenta uma abordagem para recuperação de imagem baseada em textura, por meio da medida da divergência KL entre as distribuições marginais (*marginal distributions*) das magnitudes de coeficientes *wavelet* complexas. Foi aplicado o *Dual-Tree Complex Wavelet do Kingsbury* para a decomposição da imagem e modelagem dos detalhes das magnitudes dos coeficientes da sub-banda, por meio de qualquer dois parâmetros *Weibull* ou distribuições *Gamma* para que forneça soluções de forma fechada para a divergência KL.

O trabalho de (KWITT; UHL, 2008) fornece uma maneira de medir a similaridade com um baixo custo computacional. E é baseado na contribuição de (DO; VETTERLI, 2000), no qual os autores propuseram um *framework* estatístico, na qual a etapa de extração de característica e a medida de similaridade são intimamente relacionadas entre si. Em (DO; VETTERLI, 2000), propõem medir a similaridade da imagem calculando a divergência KL entre as distribuições marginais de coeficientes da sub-banda do *Discrete Wavelet Transform* (DWT). Já em (KWITT; UHL, 2008) trabalhou-se com o *Dual-Tree Complex Wavelet Transform* (DT-CWT), originalmente proposto por (KINGSBURY, 1998), que elimina as desvantagens do DWT das dependências de deslocamento, e a falta de seletividade direcional no custo de muita redundância limitada. A principal contribuição de (KWITT; UHL, 2008) foi propor um modelo para a distribuição marginal de coeficientes de sub-banda detalhada de DT-CWT e, integrar estes modelos na estrutura do *framework* de recuperação de imagem de textura estatística de (DO; VETTERLI, 2000).

No processo de recuperação de imagem probabilística, primeiramente foi estabelecido o quadro formal de recuperação de imagens probabilística de fusão dos trabalhos de (DO; VETTERLI, 2000) e (VASCONCELOS; LIPPMAN, 2000). Suponha que tenha N imagens I_i, $1 \leq i \leq N$ no banco de dados. Cada imagem é representada por um vetor de dados $\mathbf{x}_i = \{x_{i1}, ..., x_{in}\}$, que é um elemento de algum espaço de característica $\mathcal{X} \subset \mathbb{R}^n$ e é obtido pela extração de característica (*feature extraction* (FE)). A tarefa de recuperação é buscar as W imagens mais similares para uma dada imagem I_q, de acordo com algum critério de similaridade. Assume-se que todas as imagens têm a probabilidade igual e a imagem de consulta é representada pelo vetor de dados $\mathbf{x}_q$. Do ponto de vista probabilístico, cada vetor de dados contém n realizações de variáveis aleatórias independentes e identicamente distribuídas (*independent and identically distributed* (idd)) $X_1, ..., X_n$, que segue uma distribuição paramétrica com a função de probabilidade de densidade (*Probability Density Function* (PDF)) $p(x) \mid \theta$, $\theta \in \mathbb{R}^d$. Dado um estimador consistente $\hat{\theta}$ para o vetor de parâmetro θ, pode-se utilizar $\hat{\theta}$ sem limitações. Sob estas premissas, é natural selecionar a imagem mais similar I_r para I_q, cujo parâmetro θ_r do vetor conduz para uma maximização de uma função de probabilidade/log-probabilidade

(*likelihood/log-likelihood*), por exemplo

$$r = \underset{j}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} \log(p(x_{ji} \mid \theta_j)). \tag{37}$$

Note que o fator adicional $1/n$ não afeta o resultado da maximização. Por aplicar a *weak law large numbers* para Eq. (37) que $n \to \infty$ (caso assintótico), é obtido

$$r \overset{n\to\infty}{=} \underset{i}{\operatorname{argmax}}\, \mathbb{E}_{p(x|\theta_q)}(\log(p(x \mid \theta_i))) \tag{38}$$

$$= \underset{i}{\operatorname{argmax}} \int_D p(x \mid \theta_q) \log(p(x \mid \theta_p)) dx, \tag{39}$$

onde o termo $\mathbb{E}_{p(x|\theta_q)}(\cdot)$ denota a expectativa em relação a $p(x \mid \theta_q)$ e D denota o domínio de $p(x \mid \cdot)$. Ao observar que $p(x \mid \theta_q)$ é um termo independente para maximização, pode ser reescrita a Eq. (39) como a seguinte minimização do problema:

$$r = \underset{i}{\operatorname{argmin}} \left\{ - \int_D p(x \mid \theta_q) \log(p(x \mid \theta_i)) dx \right\} \tag{40}$$

$$\equiv \underset{i}{\operatorname{argmin}} \int_D p(x \mid \theta_q) \log \left(\frac{p(x \mid \theta_q)}{p(x \mid \theta_i)} \right) dx \tag{41}$$

No entanto, o último termo na Eq. (41) é a divergência KL entre $p(x \mid \theta_q)$ e $p(x \mid \theta_i)$, que denotaria como $\mathrm{KL}(p_q \| p_i)$ usando a abreviação $p_i := p(x \mid \theta_i)$. Por isso, tem estabelecido a conexão, em que o caso assintótico ($n \to \infty$) da seleção de probabilidade máxima é equivalente a minimização da divergência KL. Para obter a segunda imagem mais semelhante para I_q, simplesmente repete o procedimento de seleção no $W - 1$ amostras restantes de imagem. Depois de W iterações tem-se uma ordenação do banco de dados de imagem induzido pela divergência KL, com a ordem relativamente definida como $I_i \leq I_j :\Leftrightarrow \mathrm{KL}(p_q \| p_i) \leq \mathrm{KL}(p_q \| p_j)$ em relação a imagem de consulta I_q.

Foi utilizado o *Dual-Tree Complex Wavelet Transform* (DT-CWT) (KINGSBURY, 1998) para computar uma representação de imagem redundante com seis sub-bandas orientadas de detalhes complexos em cada nível de decomposição. As vantagens deste *complex wavelet transform variant* são suas aproximações invariantes ao deslocamento, sua seletividade direcional e o próprio esquema de implementação eficiente por quatro paralelos 2-D DWT's. Todas estas propriedades têm o custo baixo de quatro vezes a redundância em 2-D.

Para modelar a densidade marginal das magnitudes de coeficientes complexos de subbanda de detalhe foram considerados três modelos de distribuição: a distribuição Rayleigh, os dois parâmetros do modelo Weibull e dois parâmetros da distribuição Gamma.

O PDF da distribuição Rayleigh é dada por

$$p(x \mid b) = \frac{x}{b^2} \exp\left(-\frac{x^2}{2b^2} \right), x > 0, \tag{42}$$

com o parâmetro escalar $b > 0$. Ao inserir a Eq. (42) na Eq. (41) pode-se obter uma solução de forma fechada para a divergência KL entre duas distribuições Rayleigh com parâmetros b_i e b_j como

$$\underset{\text{Rayleigh}}{\text{KL}}(p_i \| p_j) = \frac{b_i^2}{b_j^2} - 2\log(b_i) + 2\log(b_j) - 1. \tag{43}$$

A expressão na Eq. (43) envolve apenas os parâmetros de distribuição, que permite uma computação rápida em casos que têm estimadores para b_i e b_j. No decorrer do trabalho de (KWITT; UHL, 2008), usou-se *maximum-likelihood estimators* (MLE) de b cuja definição pode ser encontrado em (KRISHNAMOORTHY, 2006). No trabalho de (KRISHNAMOORTHY, 2006) os autores propuseram dois parâmetros de distribuição Weibull como uma alternativa, uma vez que inclui a distribuição Rayleigh como um caso especial e permite mais liberdade na forma, devido a um parâmetro adicional. A função de densidade de probabilidade (PDF) Weibull é dada por

$$p(x \mid c, b) = \frac{c}{b}\left(\frac{x}{b}\right)^{c-1} \exp\left\{-\left(\frac{x}{b}\right)^{c}\right\}, x > 0, \tag{44}$$

com o parâmetro de forma $c > 0$ e parâmetro de escala $b > 0$. Para $c = 2$ e $b = \sqrt{2b}$ é trivial notar que Eq. (44) reduz a distribuição Rayleigh. Soluções para o MLE's de b e c são novamente dado em (KRISHNAMOORTHY, 2006). Infelizmente, o MLE para c não tem forma explícita e pode apenas ser computada como a solução para uma equação não linear. Por meio da inserção da Eq. (44) na Eq. (41), pode-se derivar uma solução da forma fechada para a divergência KL entre duas distribuições Weibull com parâmetros c_i, b_i e c_j, b_j como

$$\underset{\text{Weibull}}{\text{KL}}(p_i \| p_j) = \Gamma\left(\frac{c_j}{c_i} + 1\right)\left(\frac{b_i}{b_j}\right)^{c_j} + \log(b_i^{-c_i} c_i) - \log\left(b_j^{-c_j} c_j\right) + \log(b_i)c_i - \log(b_i)c_j + \frac{\gamma c_j}{c_i} - \gamma - 1, \tag{45}$$

onde γ denota a negativa da função *digamma* $\psi(x) = \Gamma'/\Gamma(x)$ em $x = 1(\gamma \approx 0.577)$.

A terceira alternativa do modelo foi considerar a distribuição Gamma, que tem já sido proposta como uma alternativa para a distribuição Rayleigh para modelar as magnitudes do filtro de saída *Gabor* (MATHIASSEN; SKAVHAUG; Bø, 2002). O PDF Gamma é dado por

$$p(x \mid a, b) = \frac{b^{-a} x^{a-1}}{\Gamma(a)} \exp\left(-\frac{x}{b}\right) \tag{46}$$

Conforme (MATHIASSEN; SKAVHAUG; Bø, 2002), uma solução de forma fechada para divergência KL entre duas distribuições Gamma com parâmetros a_i, b_i e a_j, b_j existe e pode ser computadas por

$$\underset{\text{Gamma}}{\text{KL}}(p_i \| p_j) = \psi(a_i)(a_i - a_j) - a_i + \log\left(\frac{\Gamma(a_j)}{\Gamma(a_i)}\right) + a_j \log\left(\frac{b_j}{b_i}\right) + \frac{a_i b_i}{b_j}, \tag{47}$$

com MLE para a forma e para o parâmetro dado em (KRISHNAMOORTHY, 2006).

Assumindo a independência dos dados da sub-banda permite derivar uma medida simples de similaridade entre duas imagens no *framework* (DO; VETTERLI, 2000). Desde que a divergência KL possa ser expressa em termos da entropia e *cross-entropy*, pode-se aplicar a *chain rule of entropy* (COVER; THOMAS, 1991) e obter o resultado que a divergência KL *overall* entre todas sub-bandas é simplesmente a soma sobre as divergências KL individuais. Dado que $\hat{p}^I_{sk} := p(x \mid \hat{\theta}^I_{sk})$ e $\hat{p}^J_{sk} := p(x \mid \hat{\theta}^J_{sk})$ denota a distribuição ajustada para cada sub-banda de detalhe de DT-CWT das imagens I e J, a medida de similaridade final pode ser escrita como

$$S(I,J) = \sum_{s=1}^{T}\sum_{k=1}^{6} \mathrm{KL}(\hat{p}^I_{sk} \| \hat{p}^J_{sk}) \tag{48}$$

onde T denota a profundidade da decomposição de *DT-CWT*. A Eq. (48) é muito simples calcular, desde que as expressões da divergência KL apenas envolvam as estimativas dos parâmetros.

Os experimentos realizados no trabalho de (KWITT; UHL, 2008) foram comparados com os do trabalho de (DO; VETTERLI, 2000), as configurações dos experimentos foram semelhantes de ambos. Foram utilizadas as mesmas 40 imagens de textura do banco MIT *Vision Texture*[4] e dividiu-se cada imagem em 16 sub-imagens não sobrepostas. Cada sub-imagem foi normalizada pela subtração da média e dividida pelo desvio padrão, também foi conduzido um aumento do contraste na etapa de utilizar a equalização do histograma adaptativo (ZUIDERVELD, 1994). Quanto aos conjuntos de filtros para a transformada *wavelet*, utilizou-se 8-*tap* dos filtros Daubechies para o DWT e Q-Shift de Kingsbury (14,14)-*tap* filtros (níveis ≥ 2) com (13, 19)-*tap* filtros *near-orthogonal* (nível 1) para o DT-CWT (KINGSBURY, 2001).

Para avaliar o sistema de recuperação, analisou-se o número de imagens recuperadas corretamente nas top W correspondentes. Imagens recuperadas corretamente significaram que a sub-imagem faz parte da imagem de textura correspondente. Para cada imagem (por exemplo, cada sub-imagem) sabe-se que a associação correta do conjunto de índices $Q = \{r_1, ..., r_B\}$, onde B denota o número de sub-imagens. Dado aquele conjunto de índice para os tops W correspondentes é denotado por $\{q_1, ..., q_W\}$, calcula-se

$$s_k = \frac{1}{W}\sum_{i=1}^{W} 1_Q(q_i), \qquad 1_Q(x) := \begin{cases} 1, & \text{se} \quad x \in Q \\ 0, & \text{caso contrário} \end{cases} \tag{49}$$

que é dada a porcentagem das imagens recuperadas corretamente. Desde que cada imagem é subdividida em 16 sub-imagens, define-se $W = B = 16$. Também foi avaliado a performance da recuperação de acordo com o número de imagens recuperadas consideradas. Isto significa, que calculou-se a Eq. (49) para valores variados de W.

[4] http://vismod.www.media.mit.edu

O modelo Weibull conduz à forma consistente de elevadas taxas de recuperação, sendo superior ao o modelo Gamma. Em geral, os melhores resultados dos experimentos foram obtidos com a utilização do Weibull.

Concluindo o trabalho de (KWITT; UHL, 2008), a ideia central é medir a similaridade de imagem pela divergência KL entre as distribuições marginais de coeficientes de sub-banda de detalhes *wavelet*. Mostrou-se que as magnitudes de coeficientes complexos podem ser modelados por qualquer um dos dois parâmetros: Gamma ou Weibull. Com os testes foi possível validar que a distribuição Rayleigh não é flexível o suficiente para modelar estes coeficientes de sub-bandas. Também foi mostrada a superioridade do DT-CWT sobre o clássico DWT com base nos bons resultados obtidos.

Por fim, os trabalhos aqui apresentados, e maioria dos trabalhos citados presentes na literatura priorizam a melhoria da velocidade da recuperação utilizando as divergências de Bregman. Entretanto, nenhum destes trabalhos apresentam tratamentos, de forma detalhada, sobre como utilizar as divergências como funções de similaridade. Neste trabalho apresentamos métodos de tratamentos para as funções KL e GID considerando que as coordenadas dos vetores são iguais a 0, já que a função logarítmica não está definida para este valor. Esta proposta está detalhada no capítulo a seguir.

Capítulo 5

Proposta

Existem diversos métodos para caracterização de imagens, neste trabalho utilizou-se quatro abordagens que são BoVW, BSM, FISM e BoVW-SPM. Estes métodos geram histogramas, que são a representação da imagem por um vetor de característica $\mathbf{x} = (x_1, x_2, ..., x_d)$. O vetor $\mathbf{x}$ pode assumir zero em algumas coordenadas. Observa-se que a KL e a GID (Tabela 2) são definidas utilizando a função logarítmica cujo domínio é $x > 0$. Então tornou-se necessário um tratamento para os dados a fim de assegurar o domínio de definição das funções logarítmicas[1]. Na seção 5.1 será abordado alguns tratamentos que podem ser adotadas para este fim.

5.1 Tratamentos para Kullback Leibler e *Generalized I-Divergence*

Foram propostos três tratamentos para divergência KL a fim de garantir o domínio *d-Simplex*, sendo que um deles foi proposto na Teoria da Informação (TI) e, os outros dois tratamentos foram criados e analisados neste trabalho. Também foi apresentada a análise para a GID.

As características das imagens de consulta e do banco de dados foram representadas por vetores (histogramas) que podem assumir valores iguais a zero em suas coordenadas. Por outro lado, as divergências KL e GID são definidas utilizando a função logarítmica cujo domínio é $x > 0$. Desta forma, a aplicação das divergências exigiu um tratamento dos valores das componentes do vetor de característica a fim de assegurar o domínio de definição das funções logarítmicas.

A Kullback Leibler (KL) entre dois vetores $\mathbf{x}$ e $\mathbf{y}$, $\mathbf{x} = (x_1, ..., x_d)$ e $\mathbf{y} = (y_1, ..., y_d)$ é definida por:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{d} x_j \log_2 \left(\frac{x_j}{y_j} \right)$$

[1] Este estudo foi realizado conjuntamente com a Daniela Portes Leal Ferreira.

onde $\mathbf{x}$ e $\mathbf{y}$ pertencem ao conjunto *d-Simplex* (Tabela 2),

$$\triangle^d = \left\{ \mathbf{x} = (x_1, x_2, ..., x_d); 0 \leq x_j \leq 1, \text{tal que } \sum_{j=1}^{d} x_j = 1 \right\}. \tag{50}$$

Temos que, se para algum j, $x_j = 0$ e $y_j \neq 0$, a parcela $0log_2(0/y_j)$ pode ser considerada nula, visto que o limite $\lim_{x_j \to 0} x_j log_2 x_j = 0$. 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 se justifica por continuidade em (COVER; THOMAS, 1991). Assim,

$$x_j log_2 \left(\frac{x_j}{y_j} \right) = \begin{cases} 0, & se \quad x_j = 0 \\ \infty, & se \quad x_j \neq 0 \quad e \quad y_j = 0 \end{cases} \tag{51}$$

Da Eq. (51) tem-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 $x_j log_2(x_j/0) = \gamma$, onde γ é um valor suficientemente grande. Este tratamento será referido como **Kullback Leibler baseado na Teoria da Informação** (KL/TI). Outra maneira é considerar $y_j \to 0$, assumindo $y_j = \varepsilon$, onde ε é um valor bem próximo de zero. Este método será referido como **Kullback Leibler com o tratamento do ε (KL/ε)**.

Uma outra forma de garantir o domínio da função logarítmica é realizar um deslocamento das coordenadas dos vetores $\mathbf{x}$ e $\mathbf{y}$ evitando que estas assumam valores iguais a zero. Este deslocamento é obtido somando um valor $\alpha \in \mathbb{R}$ em cada coordenada dos vetores, ou seja, $\mathbf{x} = (x_1 + \alpha, x_2 + \alpha, ..., x_d + \alpha)$ e $\mathbf{y} = (y_1 + \alpha, y_2 + \alpha, ..., y_d + \alpha)$. Entretanto, os vetores $\mathbf{x}$ e $\mathbf{y}$ devem pertencer ao conjunto de domínio *d-Simplex* assim é necessário garantir que a soma das coordenadas dos vetores sejam iguais a 1. Observe como $\sum_{j=1}^{d} x_j = 1$, tomando $\alpha = \frac{1}{d}$, temos:

$$\begin{aligned} \sum_{j=1}^{d} (x_j + \tfrac{1}{d}) &= (\sum_{j=1}^{d} x_j + \sum_{j=1}^{d} \tfrac{1}{d}) \\ &= (\sum_{j=1}^{d} x_j + \tfrac{1}{d} \sum_{j=1}^{d} 1) \\ &= (1 + \tfrac{1}{d} d) = 2. \end{aligned} \tag{52}$$

Assim, sugere-se que 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 vamos considerar as coordenadas de $\mathbf{x}$, modificadas por $\frac{1}{\beta}((\beta - 1)x_j + \frac{1}{d})$ onde β é uma constante qualquer maior que 1, isto é: $\beta > 1$. A seguir é mostrado que com essa modificação dos vetores, os dados continuam no *d-Simplex*.

Proposição: A transformação afim $T(x) = \frac{1}{\beta}((\beta - 1)x + \frac{1}{d})$ leva vetores de $\triangle^d$ em $\triangle^d$.

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

$$\begin{aligned}
\sum_{j=1}^{d} \tfrac{1}{\beta}((\beta-1)x_j + \tfrac{1}{d}) &= \tfrac{1}{\beta}\sum_{j=1}^{d}((\beta-1)x_j + \tfrac{1}{d}) \\
&= \tfrac{1}{\beta}(\sum_{j=1}^{d}(\beta-1)x_j + \sum_{j=1}^{d}\tfrac{1}{d}) \\
&= \tfrac{1}{\beta}((\beta-1)\sum_{j=1}^{d} x_j + \tfrac{1}{d}\sum_{j=1}^{d} 1) \\
&= \tfrac{1}{\beta}((\beta-1)1 + \tfrac{1}{d}d) \\
&= \tfrac{1}{\beta}(\beta - 1 + 1) = 1.
\end{aligned} \tag{53}$$

Nota-se que quanto maior o valor de β, menor será o deslocamento realizado na coordenada x_j. Este deslocamento utilizando a ponderação β será referido como **Kullback Leibler com Deslocamento Normalizado** (KL/N).

No uso da função de Bregman *Generalized I-Divergence* (GID) cujo domínio é dos números reais positivos, não se faz nenhuma normalização uma vez que os vetores não estão em $\triangle^d$, lembrando aqui que a GID é uma generalização da KL em domínios mais abrangentes. A única preocupação se refere ao fato da positividade do vetor de característica.

Uma forma de evitar os zeros nas coordenadas dos vetores $\mathbf{x}$ e $\mathbf{y}$, de forma análoga ao realizado para a divergência KL, é fazer um deslocamento α de modo que translade as coordenadas de acordo com os dados do problema, ou seja, $\mathbf{x} = (x_1+\alpha, x_2+\alpha, ..., x_d+\alpha)$ e $\mathbf{y} = (y_1+\alpha, y_2+\alpha, ..., y_d+\alpha)$, não sendo necessária normalização dos vetores (como apresentado para a KL/N). Este tratamento será denominado ***Generalized I-Divergence* com o tratamento do deslocamento (GID/D)**. Também pode-se utilizar a **GID com o tratamento ε (GID/ε)**, quando os dados não estão normalizados.

Esta análise do domínio de definição das divergências, bem como os resultados dos experimentos apresentados no Capítulo 7, foram parcialmente apresentados em (ROCHA et al., 2014), e uma análise mais detalhada está sendo elaborada para publicação.

Capítulo 6

Metodologia dos Experimentos

Para a análise de recuperação de imagens utilizando as divergências de Bregman em sistemas CBIR, optou-se por quatro abordagens diferentes: BoVW, BSM, FISM e BoVW-SPM para a caracterização das imagens. A partir dessas abordagens, foi possível gerar os histogramas, que são compostos pelas informações que representam as imagens baseadas em seu conteúdo. Após a geração de histogramas foi realizada a fase de similaridade, onde fizemos o estudo das divergências de Bregman e das distâncias tradicionais, analisando as vantagens e desvantagens em cada abordagem. Apresentamos na Seção 6.1, um sumário de notações utilizadas no decorrer deste capítulo para uma melhor compreensão da descrição das técnicas e métodos. A seguir, detalhamos a metologia empregada do presente estudo, bem como suas etapas e respectivas técnicas utilizadas para realizar os experimentos.

6.1 Sumário de Notações Utilizadas

Ao decorrer deste capítulo serão utilizadas as seguintes notações:

I_{rgb} – Imagem no formato RGB.

I_{tc} – Imagem em tons de cinza.

h_{tc} - Histograma quantizado da imagem I_{tc}.

I – Representação geral de uma imagem do banco, como também a representação de uma imagem equalizada.

τ_j – Quantidade de descritores da imagem I_j.

$s_j = \{s^{(1)}, s^{(2)}, ..., s^{(\tau_j)}\}$ – Conjunto de descritores SIFT's da imagem I_j.

$S = \{s_1, s_2, ..., s_n\}$ – Conjunto de todos os descritores SIFT's do banco de dados de imagens.

$C = \{c_1, c_2, ..., c_\upsilon\}$ – Conjunto de classes, onde todos elementos de C são destintos dois a dois.

υ – Número de classes.

$\{I_1, I_2, ..., I_n\}$ – Representa o conjunto de imagens do banco de dados.

c_i – Uma classe i.

t_i – Número de imagens da classe c_i.

$n = \sum_{i=1}^{\upsilon} t_i$ – Quantidade de imagens que contém o banco de dados.

d – Dimensão (do vetor histograma)

$q_i, p, q, \mathbf{x}, \mathbf{y}$ – Representação da imagem (vetor, histograma).

$\bar{q}$ – Media aritmética.

$\ddot{q}$ – Mediana.

$\mathbb{Z}$ – Conjunto dos naturais.

I_c – Imagem de consulta.

h_c – Histograma de consulta.

h_i – Histograma que representa a imagem I_i.

$(\mathring{F}_1, \mathring{F}_2, ..., \mathring{F}_n)$ – Resultados da função de similaridade ordenado de forma crescente, quando comparado um h_c com cada h_i.

F_s – Função de similaridade.

L_I – Lista ranqueada de imagens do banco de dados.

$\{I_1, I_2, ..., I_n\}$ – Todas as imagens do banco de dados, sendo que todos os elementos são distintos dois a dois.

$L_I = (I_{\mathring{F}_1}, I_{\mathring{F}_2}, ..., I_{\mathring{F}_k})$ – Lista das k primeiras imagens ranqueada de acordo com a imagem de consulta.

6.2 Visão Geral

Para a análise de recuperação de imagens utilizando as divergências de Bregman, primeiramente foi desenvolvido um sistema de recuperação de imagens que agrega quatro abordagens diferentes: BoVW, BSM, FISM e BoVW-SPM. Cada abordagem criada gera um histograma diferente, que contém informações da imagem de acordo com seu conteúdo.

As abordagens (BoVW, BSM e FISM) utilizadas para o sistema foram replicadas de acordo com o trabalho de (SOARES; SILVA; GULIATO, 2012). A abordagem BoVW-SPM foi criada a partir da biblioteca VLFeat[1]. Todas abordagens (BoVW, BSM, FISM e BoVW-SPM) tiveram como propósito a criação das representações (histogramas) das imagens. Após a geração dos histogramas, foi calculada a similaridade com diferentes métricas e comparados os resultados no Capítulo 7.

O BoVW foi escolhido por ser uma técnica robusta, simples, eficiente e invariante a iluminação, visualização, rotação e oclusão (CSURKA et al., 2004). As demais abordagens (BSM, FISM, BoVW-SPM) são adaptações do BoVW. Deste modo, o objetivo foi comparar o desempenho dos sistemas utilizando as divergências (KL e GID), considerando estas abordagens, com sistemas que utilizam funções de similaridades convencionais.

[1] http://www.vlfeat.org/.

O sistema de recuperação de imagens criado neste estudo pode ser dividido em duas etapas: *off-line* e *on-line*. Na etapa *off-line*, como o próprio nome diz, foram realizados os procedimentos preliminares para o funcionamento do sistema. Enquanto que, na etapa *on-line*, o usuário interage com o sistema, por exemplo, escolhendo as imagens que serão a consulta. A Figura 21 mostra uma visão geral da proposta.

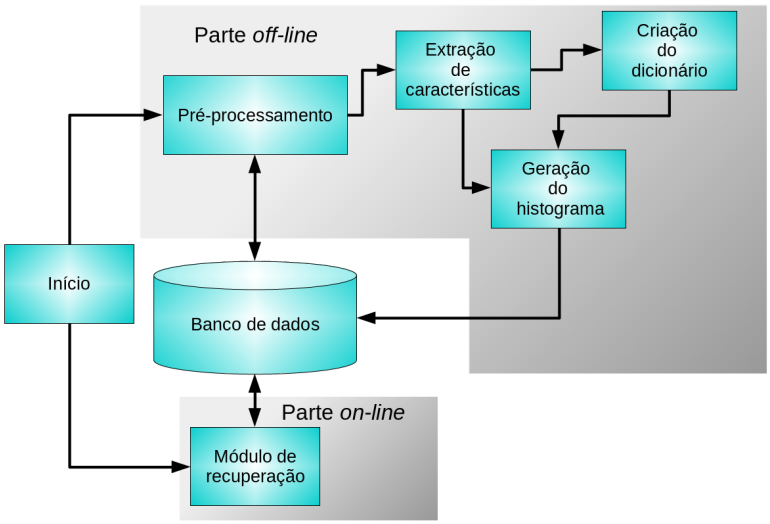

Figura 21 – Fluxograma que representa um panorama do modelo adotado.

Na etapa *off-line*, foram realizados quatro procedimentos com as imagens contidas no banco de dados, que são: pré-processamento, extração de característica, criação do dicionário de palavras visuais (também chamado de *codebook* ou apenas dicionário) e a geração do histograma. Lembrando que, para criação do histograma foi necessário obter as informações do módulo de extração de características e do módulo de criação do dicionário de palavras visuais.

Na etapa *on-line*, que foi basicamente formada por um módulo do CBIR, o usuário seleciona as imagens (que podem estar ou não no banco de dados) como entrada para o sistema e as coloca dentro de uma pasta específica do sistema, sendo que, a cada imagem de consulta é obtido um histograma correspondente, igualmente apresentado na etapa *off-line*. Em seguida, é computada a similaridade entre o histograma de consulta com os histogramas das imagens do banco de dados. Ao final do processo, de acordo com o resultado da medida de similaridade entre os histogramas, são retornadas ao usuário um *ranking* de imagens do banco, que correspondem aos histogramas do banco, mais similares de acordo com o histograma de consulta.

6.3 Parte *off-line* do Sistema

6.3.1 Pré-Processamento

O pré-processamento é a primeira etapa do sistema, na qual os dados foram preparados para serem posteriormente processados para os outros módulos segundo demonstrado na Figura 21. As etapas do pré-processamento foram: (I) transformação das I_{rgb} para imagens em tons de cinza I_{tc}, (II) equalização da I_{tc} de acordo com seu histograma de frequência e, por fim, é obtido o mapa de saliência (MS) (apresentado na subseção 3.2.1). De acordo com o MS proposto por (ITTI; KOCH; NIEBUR, 1998), foi feita a binarização (III) da I_{tc} (chamado aqui de mapa de saliência binário) e utilizar a Teoria dos Conjuntos Nebulosos (TCN)(IV).

As imagens I_{rgb}'s do banco de dados foram transformadas em I_{tc}'s. E, em seguida, as I_{tc}'s foram quantizadas em 256 níveis de cinza criando um histograma correspondente h_{tc} – como mostrado na Figura 2 da subseção 2.2.1. O h_{tc} gerado foi equalizado para obter assim uma imagem I com o melhor contraste[2]. Com a imagem equalizada I foi obtido o MS pelo modelo de Itti apresentado na subseção 3.2.1.

A abordagem BoVW é composta pelas etapas (I), (II), descritas acima. Enquanto que, a abordagem BSM utilizou as etapas (I), (II) e (III). O FISM foi formado pelas etapas (I), (II), (IV). Por fim, a abordagem BoVW-SPM utilizou apenas a etapa (I).

6.3.2 A Extração de característica

Depois do módulo de pré-processamento, na imagem equalizada I_i foi aplicado o descritor SIFT no modo *sparse* (S-SIFT) (V) e *dense* (D-SIFT) (VI) (subseção 3.1.1). O descritor SIFT proposto por Lowe (como apresentado na subseção 3.1.1) foi aplicado para detectar os pontos de interesse da imagem I_i gerando um conjunto de SIFT's $s_i = \{s^{(1)}, s^{(2)}, ..., s^{(\tau_i)}\}$, onde τ_i é a quantidade de descritores gerados para a I_i.

Na abordagem BoVW foi utilizado o descritor S-SIFT para discriminar o valor de I. Para abordagem BSM, o S-SIFT foi combinado com a etapa (III) selecionando a região de interesse como *foreground* com o *thresholding* percentual fixo igual a 25%, como descrito na subseção 3.2.2. Para a abordagem FISM foi combinada a etapa (IV) com o S-SIFT, sendo que a região de interesse foi a *foreground*, como mostrado na subseção 3.2.3. Por fim, a abordagem BoVW-SPM utilizou apenas o descritor D-SIFT para detecção dos pontos de interesse da I_{tc}.

[2] O contraste é uma medida qualitativa que está relacionada com a distribuições dos tons de cinza em uma imagem.

6.3.3 A Criação do dicionário de palavras visuais

Ao terminar o módulo de extração de caraterísticas de todas as imagens do banco $\{I_1, I_2, ..., I_n\}$ foram construídos dois dicionários diferentes: o primeiro usou os $\{s_1, s_2, ..., s_n\}$ extraídos pelo S-SIFT's (VII) e o outro utilizou $\{s_1, s_2, ..., s_n\}$ extraído pelo D-SIFT (VIII). Os dicionários foram criados como demonstrado na seção 3.1.

Para construir o dicionário (VII) foi utilizado o algoritmo *k-means* (JAIN; MURTY; FLYNN, 1999) enquanto que para o dicionário (VIII) foi usado o algoritmo *k-means* baseado na desigualdade triangular proposto por (ELKAN, 2003).

A abordagem BoVW construiu o *codebook* (VII). As abordagens BSM e FISM usaram o mesmo dicionário criado para abordagem BoVW. Enquanto que o dicionário (VIII) foi criado para abordagem BoVW-SPM.

6.3.4 A Criação dos histogramas

Após terminar os procedimentos de extração de características das imagens e a criação do dicionário de palavras visuais (da Figura 21), criaram-se os histogramas $\{h_1, h_2, ..., h_n\}$ que representaram as $\{I_1, I_2, ..., I_n\}$ imagens do banco de dados. As características s_i de cada imagem I_i foram mapeadas para a palavra visual mais próxima do dicionário, obtendo assim um histograma de palavras visuais associadas a cada imagem contida no banco de dados, como descrito na seção 3.1.

Na abordagem BoVW, BSM e FISM, o histograma de palavras visuais h_i gerado do mapeamento das características s_i, obtidas pelo S-SIFT, com a palavra visual mais próxima do dicionário (VII). Lembrando que o s_i da abordagem BSM e FISM foram combinados com a etapa (III) e (IV) respectivamente, como mostrado nas subseções 3.2.2 e 3.2.3.

Para a abordagem BoVW-SPM, o histograma foi criado de acordo com o dicionário de palavras visuais (VIII), aplicando três níveis (0 até 2) das pirâmides espaciais (como apresentado na subseção 3.3) para representação da imagem. Após obter um histograma para cada imagem, foi realizada a normalização destes histogramas, de modo que suas coordenadas ficaram no intervalo de [0, 1], e a soma de todos os valores do histograma é igual a 1. Em sequência, apresentamos a parte *on-line* do sistema.

6.4 Parte *on-line* do Modelo

Como já dito anteriormente, no contexto de recuperação de imagens baseada em conteúdo faz-se necessária a utilização de uma medida de similaridade para comparar as características de uma imagem de consulta com as características de cada uma das imagens do banco de dados. Desta forma, teremos a comparação da imagem consulta com todas as imagens que estão contidas no banco, retornando as imagens mais similares de

acordo com a consulta. Para medir a similaridade foram implementadas funções tradicionais (Cosseno e Euclidiana – subseção 2.2.2.1) e as divergências de Bregman com seus tratamentos (KL/ε, KL/N, KL/TI, GID/D e GID/ε).

Ao realizar uma consulta, pode-se escolher uma das abordagens BoVW, BSM, FISM ou BoVW-SPM. A partir da imagem de consulta I_c, cria-se o histograma de consulta correspondente h_c, como apresentado na etapa *off-line*. Lembrando que cada abordagem cria um histograma diferente para I_c. Também é possível escolher qual função de similaridade F_s será aplicada para comparar o h_c com todos histogramas $(h_1, h_2, ..., h_n)$ da base de dados. As funções de similaridades que podem ser escolhidas são a Euclidiana, Cosseno, GID/D, GID/ε, KL/ε, KL/N, KL/TI com os seus parâmetros. Os parâmetros da GID/D, GID/ε, KL/ε, KL/N, KL/TI devem ser configurados antes de calcular a similaridade. Neste trabalho foram realizados testes empíricos para determinar o parâmetro que maximiza o resultado.

A consulta por similaridade foi realizada utilizando o método dos k-vizinhos mais próximos (apresentada na subseção 2.2.2) de acordo com a comparação da função de similaridade $F_s(h_c, h_i)$, onde $\forall\ i \in \{1, ..., n\}$ e $1 \leq k \leq n$. Os resultados das funções de similaridade são organizadas em ordem crescente $(F_s(h_c, h_l) \leq F_s(h_c, h_z) \leq ... \leq F_s(h_c, h_w)) = (\mathring{F}_1, \mathring{F}_2, ..., \mathring{F}_n)$, onde $\forall\ l, z, w \in \{1, ..., n\}$, e são associadas aos histogramas correspondentes. Sendo assim, torna-se possível escolher as k imagens que correspondem aos $(\mathring{F}_1, \mathring{F}_2, ..., \mathring{F}_k)$, que será a lista ranqueada de imagens $L_I = (I_{\mathring{F}_1}, I_{\mathring{F}_2}, ..., I_{\mathring{F}_k})$ do banco de dados que será retornada ao usuário.

Com a lista L_I são aplicados os métodos de avaliação *precision at k* (P@k) (precisão em k), MAP, Precisão x Revocação e nDCG para medir o desempenho da recuperação.

Capítulo 7

Experimentos e Análise dos Resultados

Apresentamos neste capítulo, os experimentos realizados e os resultados obtidos pelo estudo comparativos proposto, detalhando o desempenho da GID e da KL como funções de similaridade na etapa *on-line* do sistema e, demostrando a qualidade dos resultados obtidos no cenário da recuperação de imagem por conteúdo.

Os experimentos foram divididos em 3 grupos: grupo 1 (experimento I e II), grupo 2 (experimento III) e grupo 3 (experimento IV):

- Experimento I e II: No primeiro grupo de experimentos, apresentamos uma comparação dos resultados de busca de imagens utilizando as funções GID, KL, Cosseno e Euclidiana. Foram utilizados quatro bancos de dados (Caltech101[1], Oxford[2], UKbench[3] e o Holiday[4]) e quatro formas de obtenção do histograma BSM, FISM, BoVW e BoVW-SPM.

- Experimento III: Neste experimento analisamos quatro tratamentos da função $\log x$ e apresentamos os resultados obtidos. Como as funções de similaridade GID e KL utilizam a função logarítmica em sua formulação (veja a Tabela 2), os dados necessitam de tratamento, uma vez que o $\log x$ só é definida se $x > 0$, e nas caracterizações usadas, os vetores de características representam em cada coordenada uma frequência de ocorrência de palavras visuais, as quais podem ser nulas. Verificou-se que tratamentos distintos interferem nos resultados obtidos.

- Experimento IV: Os resultados desta pesquisa são comparados com dois trabalhos existentes na literatura, (XU et al., 2012) e (JÉGOU; CHUM, 2012), utilizando os mesmos bancos de dados, as mesmas consultas e os mesmos métodos de avaliação, mas com abordagens de caracterização diferentes. Os bancos de dados utilizados neste experimento foram Caltech101 e o Holiday.

[1] http://www.vision.caltech.edu/Image_Datasets/Caltech101/.
[2] http://www.robots.ox.ac.uk/ṽgg/data.html
[3] http://www.vis.uky.edu/s̃tewe/ukbench/
[4] http://lear.inrialpes.fr/j̃egou/data.php#holidays

7.1 Bancos de dados

Os bancos de dados utilizados para realização do nosso estudo são:

- **Caltech101**: Este banco de dados é considerado muito heterogêneo e de médio porte. Contém um total de 9.144 imagens de tamanhos variados. Possui 101 categorias diferentes que inclui imagens de animais, insetos, plantas, pequenos objetos, veículos, partes do corpo humano, desenhos, etc. O número de imagens por categorias variam de 31 até 800 imagens. O banco Caltech101 é considerado desafiador pois inclui imagens com alta variação intra-classe e oclusão de fundo. A Figura 22 apresenta uma amostra de algumas imagens do banco Caltech101.

- **Oxford**: Foi criado pelo grupo de geometria visual da universidade de Oxford, e é composto por 13 categorias diferentes, com um total de 9.025 imagens. Para realização deste trabalho, utilizaram-se apenas 5 classes: *Motorbikes* (826 imagens), *Airplanes* (1074), *Faces* (450), *Leaves* (186) e *Guitars* (1030), totalizando 3.566 imagens. A Figura 23 mostra uma imagem de cada uma das categorias.

- **UK-bench**: Criado pelo grupo de pesquisadores do *Center for visualization & Virtual Environments* da universidade de Kentucky, têm um total de 10.200 imagens em que cada classe do banco é composta por exatamente quatro imagens, ou seja, um total de 2.550 categorias. Todas as imagens do banco têm um tamanho padrão de 640×480. A Figura 24 ilustra um exemplo de quatro classes com as respectivas imagens que a compõem.

- **Holiday**: Criado pelos integrantes do projeto *Agence Nationale de la Recherche* RAFFUT (ANR RAFFUT). Formado por 500 categorias diferentes, com um total de 1.491 imagens. O número de imagens por categoria varia de 2 até 13 imagens. Este banco contém principalmente fotos de férias pessoais, contendo variações como: rotação, ponto de vista, mudanças de iluminação, desfocagem, etc. Contém um grande variedade de cenas (natural, artificial, efeitos de água e fogo, etc) e são imagens de alta resolução. Alguns exemplos de imagens deste banco estão ilustrados na Figura 25.

7.2 Métodos utilizados para a avaliação dos resultados

Foi utilizada a média aritmética de todas as consultas de acordo com os experimentos, aplicando as medidas de avaliação *Mean Average Precision*, *normalized Discounted Cumulative Gain* (nDCG), média de precisão em k e precisão x revocação, para calcular o desempenho das distâncias utilizadas na recuperação de imagens baseadas em conteúdo.

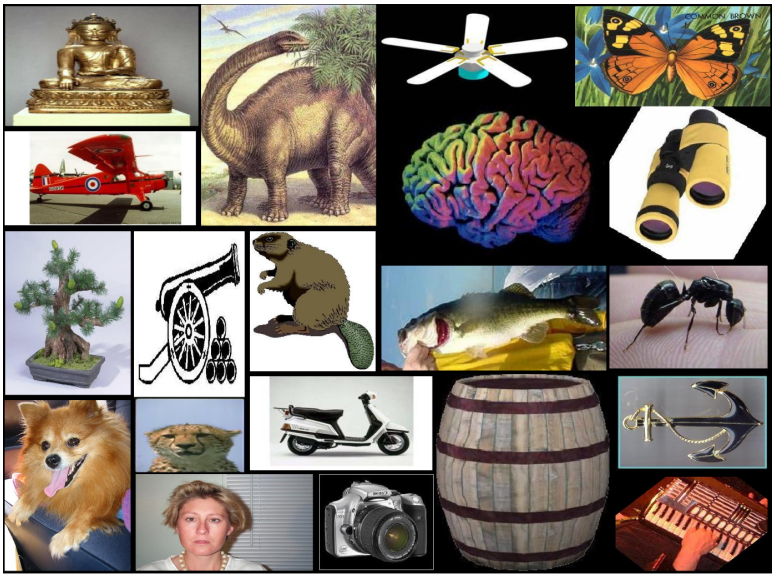

Figura 22 – Exemplo de algumas imagens que estão contidas no banco Caltech101.

Figura 23 – Cinco imagens selecionadas de forma aleatória do banco de dados Oxford, sendo uma imagem de cada classe.

Figura 24 – Exemplo das imagens que formam quatro classes diferentes do banco UK-bench.

O *Mean Average Precision* representa o resultado com 100% de revocação e será denotado de apenas MAP. O MAP@200 é a avaliação MAP com as 200 primeiras imagens da lista ranqueada. A média de precisão em k foi representada por (MP@k). E, a medida nDCG foi utilizada para as primeiras 10 (nDCG@10), 20 (nDCG@20), 30 (nDCG@30) e 100 (nDCG@100) respostas da lista ranqueada. A última medida utilizada foi a precisão

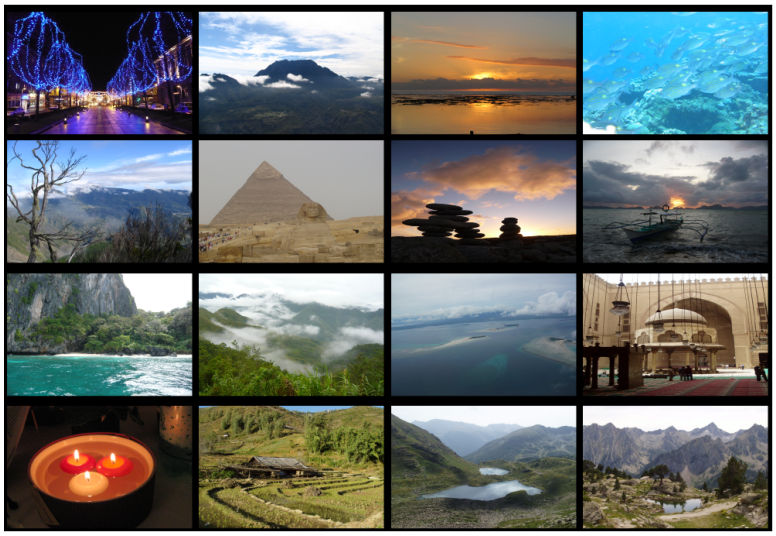

Figura 25 – Exemplo de algumas imagens que estão contidas no banco Holiday.

x revocação que mostra os resultados da precisão em diferentes níveis de revocação.

7.3 Condução dos experimentos

Os experimentos foram conduzidos de forma a validar as hipóteses apresentadas na Seção 1.3. Os experimentos descrito na subseção 7.3.1 apresenta a comparação das divergências de Bregman (KL e GID) para averiguar sua eficiência, na etapa de similaridade, quando comparadas com as distâncias tradicionais. Já a subseção 7.3.2 apresenta os experimentos e resultados com os tratamentos aplicados para assegurar o domínio das funções logarítmicas nas divergências. Para cada experimento existe um conjunto de parâmetros que serão descritos nas suas respectivas subseções.

7.3.1 Grupo 1 de experimentos

7.3.1.1 Experimentos I

O experimento I foi conduzido utilizando um subconjunto de 3.566 imagens do banco de dados Oxford (Seção 7.1) para efeito de comparação de resultados, e o banco Caltech101 para a análise das divergências. Para o banco Oxford foram utilizadas as abordagens BoVW, FISM e BSM (apresentados na seção 3.1 e subseções 3.2.2 e 3.2.3, respectivamente) e, na condução desse experimento foram utilizados os mesmos parâmetros descritos em (SOARES; SILVA; GULIATO, 2012; NAKAMOTO; TORIU, 2011).

Banco Oxford

Neste tópico são apresentados os resultados dos experimentos realizados na base de dados Oxford. Foram utilizados 10% de cada classe como consulta (as mesma utilizadas

em (SOARES; SILVA; GULIATO, 2012)), totalizando 355 buscas por similaridade. O tamanho do dicionário de palavras visuais (*codebook*) foi configurado em 1.000 grupos para todas as abordagens. O FISM e o BSM utilizaram o mapa de saliência (MS) proposto em (ITTI; KOCH; NIEBUR, 1998) e a região de interesse escolhida para ambos os métodos foi o *foreground*. O *threshold* t do método BSM (subseção 3.2.2) para separação do *foreground* foi configurado em 25%, visto ser o melhor resultado obtido em (SOARES; SILVA; GULIATO, 2012).

Neste experimento foi utilizado para a função GID, um deslocamento $\alpha \in \mathbb{R}$, $\alpha > 0$ em todos os vetores característicos tanto da consulta como para todos os elementos da base de dados. Vários valores foram testados e os resultados com $\alpha = 1$ foram os escolhidos para serem apresentados neste experimento. As Tabelas 4 e 5 apresentam os dados para os bancos Oxford e o Caltech101, respectivamente. A porcentagem positiva indica a melhora do índice de recuperação da GID comparada com a função Cosseno e a porcentagem negativa significa o contrário, ou seja, o quanto a GID teve de desempenho inferior comparado à medida Cosseno.

Em ambas as tabelas foram apresentados os resultados em termos de MAP, nDCG@10, nDCG@20, nDCG@100, MP@10, MP@20 e MP@30, considerando as funções Cosseno e GID. Vale ressaltar que não foram apresentados os resultados da recuperação utilizando a distância Euclidiana como função de similaridade entre os histogramas devido ao fato dos resultados serem inferiores quando comparados com a medida Cosseno.

Tabela 4 – Comparação dos resultados obtidos no banco Oxford com as funções Cosseno e a GID.

Fun. de Simil.	Cosseno			GID		
Abordagens	FISM	BoVW	BSM	FISM	BoVW	BSM
MP@10	0,5890	0,5735	0,5994	0,6473 +09,90%	0,7478 +30,39%	0,6988 +16,58%
MP@20	0,5490	0,5231	0,5385	0,5991 +09,13%	0,6723 +28,52%	0,6307 +17,12%
MP@30	0,5256	0,4900	0,5035	0,5769 +09,76%	0,6338 +29,35%	0,5937 +17,91%
nDCG@10	0,6422	0,6284	0,6516	0,7001 +09,02%	0,7929 +26,18%	0,7472 +14,70%
nDCG@20	0,5987	0,5776	0,5950	0,6514 +08,80%	0,7289 +26,19%	0,6873 +15,51%
nDCG@100	0,4944	0,4568	0,4631	0,5464 +10,52%	0,5747 +25,81%	0,5390 +16,39%
MAP	0,3903	0,3555	0,3179	0,3850 –01,36%	0,3719 +04,61%	0,3639 +14,47%

Observando os resultados apresentados na Tabela 4, vemos um ganho superior a 8% considerando todas as características (FISM, BoVW e BSM) quando utilizou-se a GID ao invés da Cosseno para medir a similaridade, chegando a 30% no MP@10 na abordagem BoVW. A única medida que não trouxe ganho foi o MAP na caracterização FISM, porém a porcentagem negativa é muito pequena, próximo de 1%. Analisando a caracterização BoVW, nota-se um ganho de até 30%, que é bastante relevante para o contexto de recuperação, desde que os usuários verifiquem os primeiros resultados da lista ranqueada.

Além dessas medidas, foram analisadas a precisão e a revocação médias destes resultados e os gráficos são apresentados nas Figuras 26, 27 e 28 referentes às aproximações

FISM, BoVW e BSM respectivamente. Os resultados obtidos na medidas de precisão x revocação no uso da GID apresentam uma superioridade no método BSM se comparada com o uso da função Cosseno em todos os níveis de revocação, apresentando uma perda de precisão nas taxas de revocação acima de 25% na aproximação FISM e 40% no método BoVW. Analisando a Figura 28, acreditamos que GID foi superior que a medida Cosseno, devido os histogramas gerados terem uma maior quantidade de 0's em suas coordenadas, pois a abordagem BSM considera apenas a região de *foreground* para gerar os histogramas. É interessante notar que a GID é superior em até 25% da taxa de revocação, o que é uma importante questão para a CBIR, visto que os usuários se interessam principalmente pelas primeiras respostas da lista do ranqueamento.

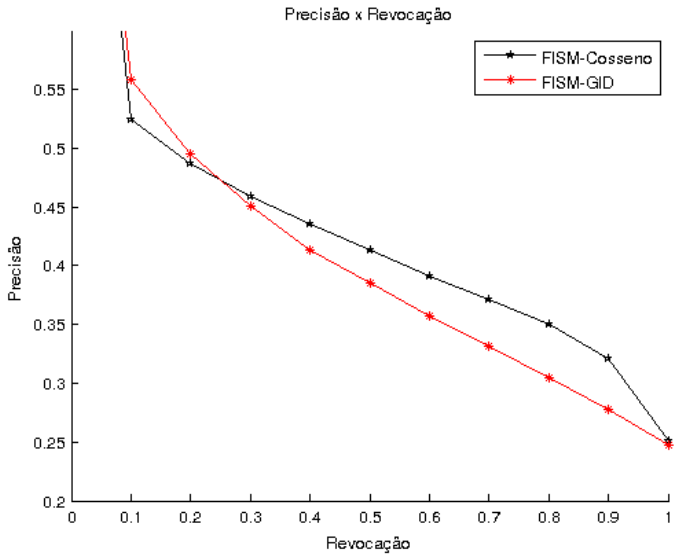

Figura 26 – Precisão x revocação média utilizando a abordagem FISM com a medida Cosseno e a GID para o cálculo de similaridade no banco Oxford.

A qualidade dos resultados são mostrados na Figura 29, o qual mostra o exemplo das 10 primeiras respostas da recuperação utilizando como medida de similaridade a Cosseno e a GID. A primeira recuperação (a) e (b) são as primeiras top 10 respostas usando a aproximação BoVW, a recuperação (c) e (d) são as primeiras 10 respostas usando a abordagem FISM e por último (f) e (g) as primeiras top 10 respostas de uma lista ranqueada utilizando o método BSM. Pode-se notar que a GID (b; d; f) é mais bem-sucedida em situações onde a função Cosseno falha.

Os resultados médio por classe da avaliação MP@10 da Tabela 4, utilizando a abordagem FISM, BoVW e BSM com as medidas Cosseno e GID, são apresentados nas Figuras 30, 31 e 32, respectivamente.

Analisando a Figura 30, observe que a Cosseno obteve melhores resultados do que o uso da função GID em 3 classes (*Faces*, *Leaves* e *Motobikes*, que são as classes 2, 4 e 5, respectivamente), enquanto a GID obteve resultados superiores apenas em duas classes. Entretanto, vale ressaltar que as avaliações inferiores alcançados pela GID conseguiu

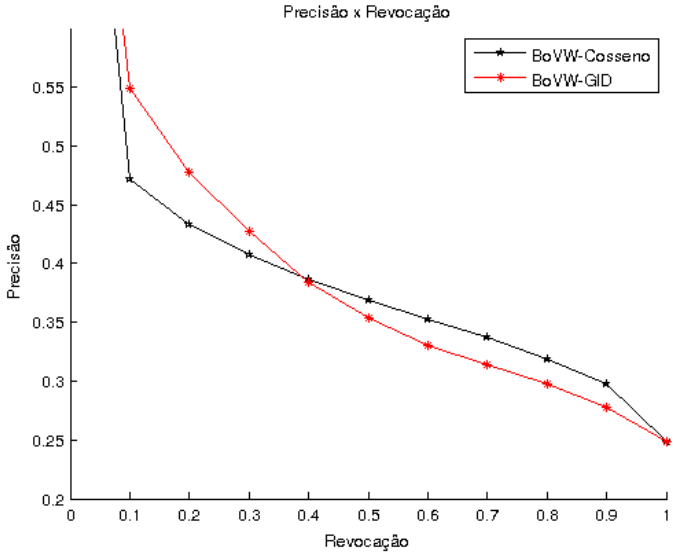

Figura 27 – Precisão x revocação média usando o método BoVW com a medida Cosseno e a GID para o cálculo de similaridade no banco Oxford.

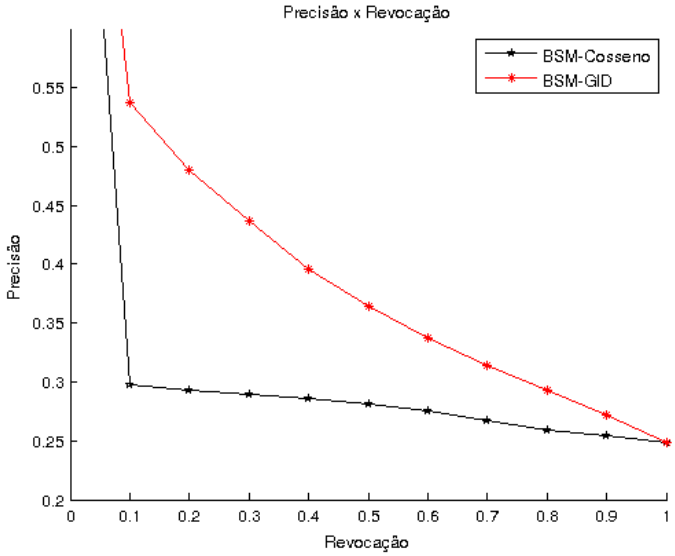

Figura 28 – Precisão x revocação média aplicando a abordagem BSM com a medida Cosseno e a GID para o cálculo de similaridade no banco Oxford.

resultados próximos da avaliação Cosseno. Já para as classes que os resultados da GID foram superiores que a Cosseno (*Airplane* e *Guitars*, que representam as classes 1 e 3, respectivamente), as diferenças entre as avaliações obtidas são maiores.

Na Figura 31 observa-se que o uso da função Cosseno, na avaliação MP@10 por classe, o maior resultado obtido foi na classe 2 (MP@10=0,8155), enquanto a menor avaliação foi na classe 3 (MP@10=0,5330). Já para a medida GID, o menor e maior resultado foram das classes 4 (MP@10=0,6166) e 2 (MP@10=0,8222), na devida ordem. Observa-se também que todas as classes utilizando a função GID para o cálculo de similaridade foram superiores do que a Cosseno, mostrando a superioridade da divergência utilizando

Figura 29 – Para cada consulta, a primeira linha mostra a recuperação utilizando a medida Cosseno (a; c; e) e a segunda linha a GID (b; d; f).

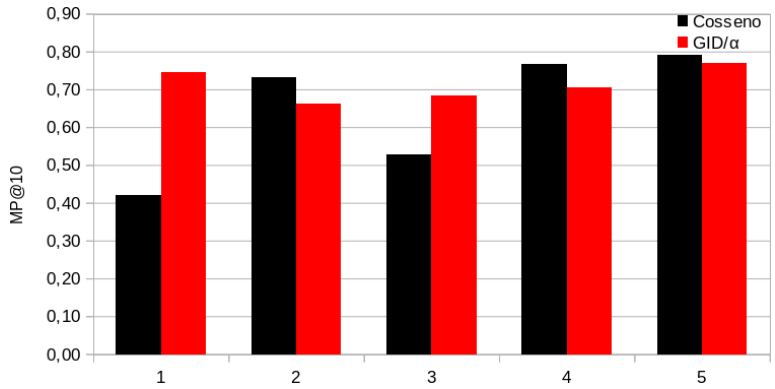

Figura 30 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Oxford com as funções de similaridade Cosseno e GID utilizando a abordagem FISM.

a abordagem BoVW.

Por fim, observando a Figura 32, nota-se que a performance da GID na avaliação por classe utilizando o MP@10 foi superior do que a função Cosseno em 4 classes (*Faces*, *Guitars*, *Leaves* e *Motobikes*). Além disso, de acordo com gráfico da Figura 32 as avaliações adquiridas com as medidas Cosseno e a GID na classe *Airplane* foram próximos, ressaltando que esta foi a única classe que a Cosseno obteve melhor resultado com a abordagem BSM.

De forma geral, a GID apresentou uma maior estabilidade e qualidade nos resultados do que a similaridade Cosseno, com o uso de diferentes abordagens (FISM, BoVW e BSM). Mostrando que a divergência de Bregman pode ser uma boa opção como função

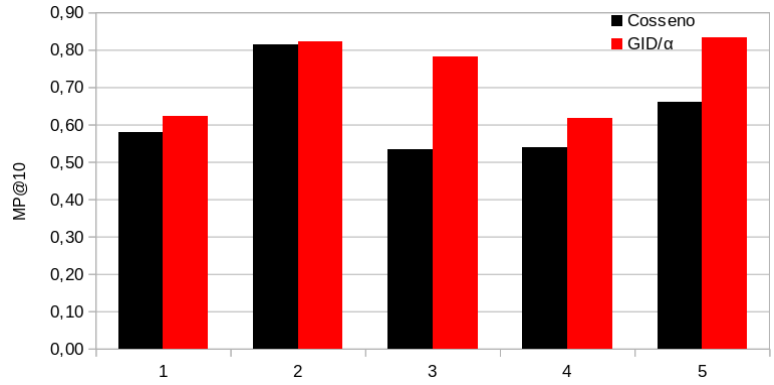

Figura 31 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Oxford com as funções de similaridade Cosseno e GID utilizando a abordagem BoVW.

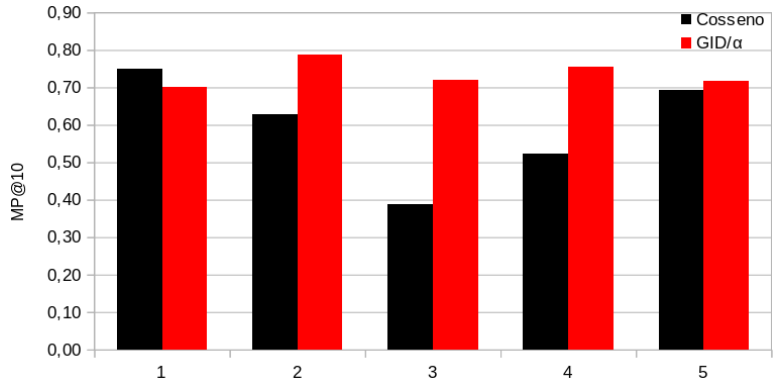

Figura 32 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Oxford com as funções de similaridade Cosseno e GID utilizando a abordagem BSM.

para medir a similaridade entre histogramas.

Banco Caltech101

Neste tópico são apresentados os resultados dos experimentos realizados na base de dados Caltech101. Para a análise das divergências no banco Caltech101, que é considerado um banco mais complexo devido o seu grande número de imagens e classes, no processo de geração dos histogramas considerou-se um procedimento semelhante ao descrito acima (tópico do Banco Oxford) nesta subseção, modificando apenas o tamanho do dicionário para 250 ao invés de 1.000, enquanto os demais parâmetros foram mantidos. Analisando-se os resultados, apresentados na Tabela 5, nota-se que o menor ganho da GID em relação à função Cosseno foi de 14%, considerando todas as abordagens para a caracterização (FISM, BoVW e BSM), e o maior ganho foi de 86% na avaliação MAP com abordagem BSM. De forma geral, os resultados mostram que a GID superou a medida Cosseno em todos os métodos de avaliação com as três diferentes abordagens de caracterização.

Tabela 5 – Comparação dos resultados obtidos no banco Caltech101 com as funções de similaridade Cosseno e a GID, com as abordagens FISM, BoVW e BSM.

Fun. de Simil.	Cosseno			GID		
Abordagens	FISM	BoVW	BSM	FISM	BoVW	BSM
MP@10	0,2340	0,2439	0,2222	0,2951 +26,11%	0,3321 +36,16%	0,3482 +56,71%
MP@20	0,1785	0,1822	0,1650	0,2315 +29,69%	0,2682 +47,20%	0,2849 +72,67%
MP@30	0,1559	0,1601	0,1416	0,2051 +31,56%	0,2423 +51,34%	0,2590 +82,91%
nDCG@10	0,3182	0,3251	0,3085	0,3778 +18,73%	0,4097 +26,02%	0,4276 +38,61%
nDCG@20	0,2543	0,2582	0,2428	0,3091 +21,55%	0,3425 +32,65%	0,3603 +48,39%
nDCG@100	0,1521	0,1593	0,1380	0,1928 +26,76%	0,2271 +42,56%	0,2406 +74,35%
MAP	0,0909	0,0890	0,0666	0,1037 +14,08%	0,1196 +34,38%	0,1240 +86,19%

As Figuras 33, 34 e 35 apresentam os gráficos de precisão e a revocação média dos resultados da Tabela 5, referentes as abordagens FISM, BoVW e BSM. Observe que a divergência de Bregman (GID) em todas as avaliações de precisão x revocação (Figuras 33, 34 e 35) foram superiores que a similaridade Cosseno em todos os níveis de revocação. Com estes resultados obtidos, a função GID novamente mostra ser mais adequada como função de similaridade entre histogramas na etapa de similaridade, com diferentes níveis de revocação e com distintas abordagens para representar as imagens, no banco Caltech101.

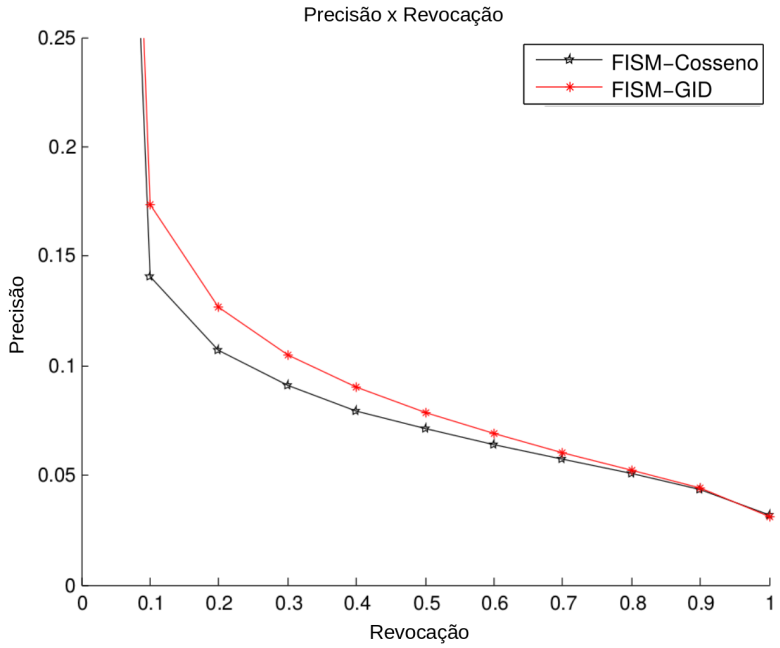

Figura 33 – Precisão x revocação média utilizando a abordagem FISM com a medida Cosseno e a GID para o cálculo de similaridade no banco Caltech101.

As Figuras 36, 37 e 38 são os gráficos do resultado médio por classe da avaliação MP@10 da Tabela 5, utilizando a abordagem FISM, BoVW e BSM com as medidas de similaridade Cosseno e GID. Analisando o gráfico da Figura 36, observa-se que de um total de 101 classes do banco Caltech101, o uso da GID conseguiu ser superior que a similaridade Cosseno em 64 classes (representando 62,75%) do banco Caltech101. Para

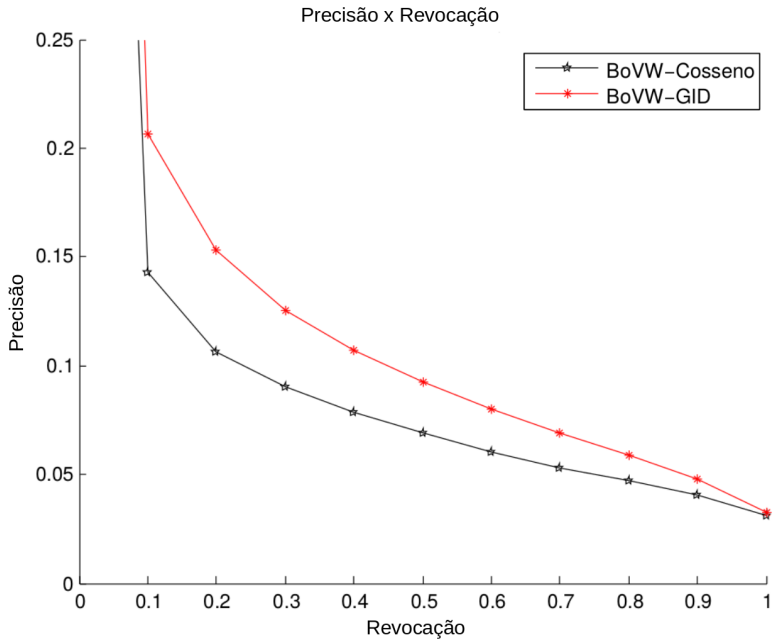

Figura 34 – Precisão x revocação média usando o método BoVW com a medida Cosseno e a GID para o cálculo de similaridade no banco Caltech101.

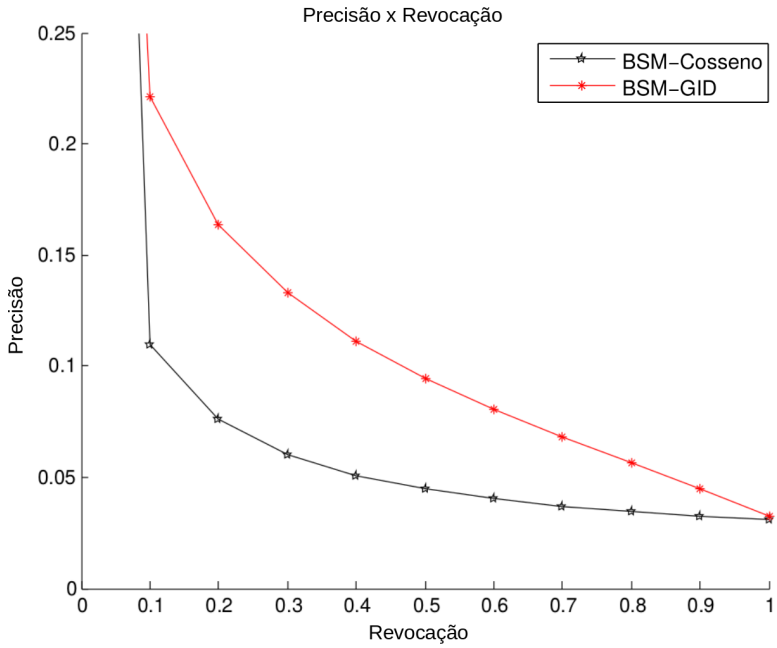

Figura 35 – Precisão x revocação média aplicando a abordagem BSM com a medida Cosseno e a GID para o cálculo de similaridade no banco Caltech101.

o gráfico da Figura 37 a medida GID obteve resultados superiores em 71,28% do total das classes do banco Caltech101, qual corresponde a 73 classes, quando comparada com a função Cosseno. Por fim, a Figura 38 apresenta a avaliação MP@10 por classe da abordagem BSM. Veja que a GID obteve avaliações superiores em 80 classes (79,21%) do banco Caltech101, enquanto a similaridade Cosseno obteve melhores resultados em apenas 21 classes. De uma forma geral, o uso da GID como função de similaridade mostrou-se

mais apropriada nas diversas classes do banco Caltech101.

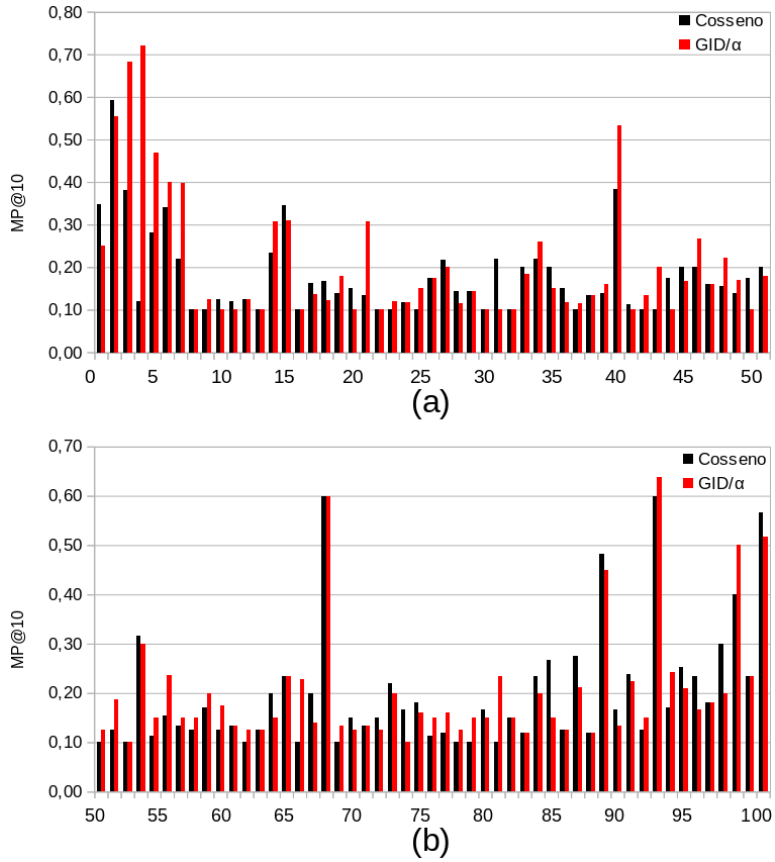

Figura 36 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Caltech101 com as funções de similaridade Cosseno e GID utilizando a abordagem FISM, sendo divididas as 101 classes em duas partes, da classe 1 até a classe 50 (a) e da classe 51 até 101 (b).

A Figura 39 apresenta um exemplo de busca onde a imagem consulta não pertence ao banco. Para isto, escolhemos 3 diferentes imagens, duas motos e um avião. Perfazendo a busca no banco de dados Caltech101, que contém imagens de motos e aviões, apresentamos as primeiras 10 imagens ranqueadas utilizando como medida de similaridade o Cosseno e a GID. Neste exemplo, a caracterização usada foi BoVW para gerar os histogramas. Na primeira moto, resultados (a) e (b), observamos que o resultado usando o Cosseno para avaliar a similaridade dos histogramas não trouxe nenhuma moto nas 10 primeiras posições do *ranking*, enquanto que utilizando a GID, temos as 10 primeiras posições retornadas com imagens de motos. A segunda moto utilizada foi uma moto amarela. Neste caso a função Cosseno apresentou apenas uma imagem significativa, ou seja, uma moto entre as 10 primeiras ranqueadas, enquanto que a GID trouxe 8 imagens relevantes. A terceira busca foi feita utilizando um avião como *query* (consulta). Aqui também temos que, enquanto a Cosseno trouxe apenas um avião nas 10 primeiras posições a GID apresentou 7 aviões entre as 10 primeiras. Este exemplo serve para evidenciar as diferenças e a superioridade das divergências para análise da similaridade em sistemas CBIR.

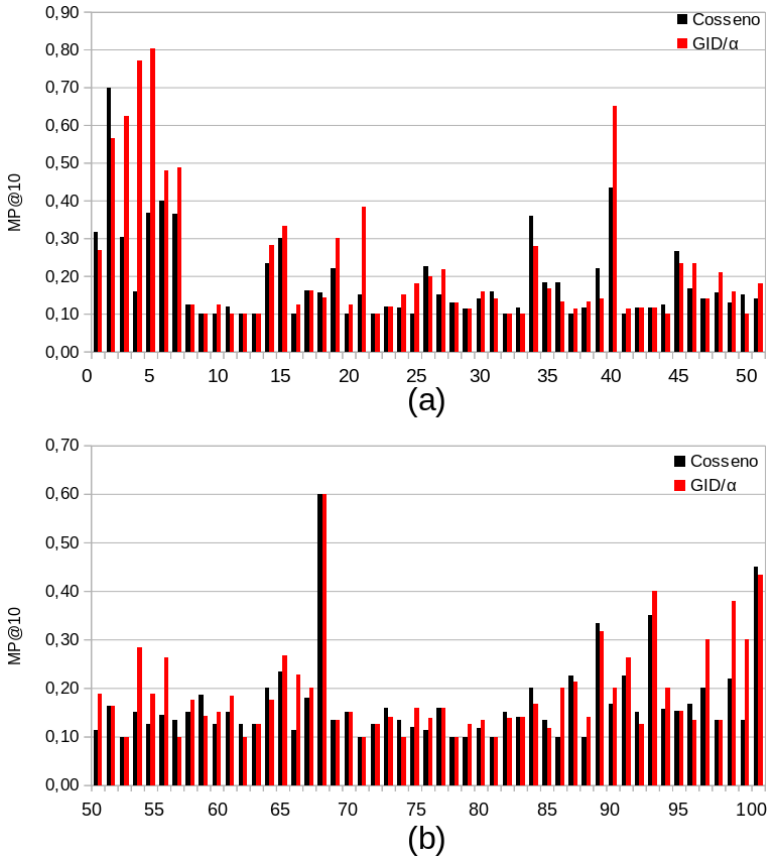

Figura 37 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Caltech101 com as funções de similaridade Cosseno e GID utilizando a abordagem BoVW, sendo divididas as 101 classes em duas partes, da classe 1 até a classe 50 (a) e da classe 51 até 101 (b).

Por fim, por meio dos resultados obtidos percebe-se que a GID é uma boa opção para quantificar a similaridade entre duas características de imagens, quando a caracterização é feita usando histogramas de frequência, comprovando também a hipótese estabelecida (apresentada na Seção 1.3), e que em muitos casos a função Cosseno não é uma boa escolha comparando se desempenho ao da GID.

7.3.1.2 Experimentos II

No experimento II, os bancos de dados foram caracterizados utilizando a abordagem BoVW-SPM da biblioteca VLFeat[5]. Os bancos utilizados foram o Caltech101, UK-Bench e o Holiday. A VLFeat é uma biblioteca de código fonte aberto no campo de visão computacional e suporta várias técnicas como SIFT, *k-means*, *hierarchical k-means* e dentre outras que estão implementadas (VEDALDI; FULKERSON, 2010).

[5] http://www.vlfeat.org/.

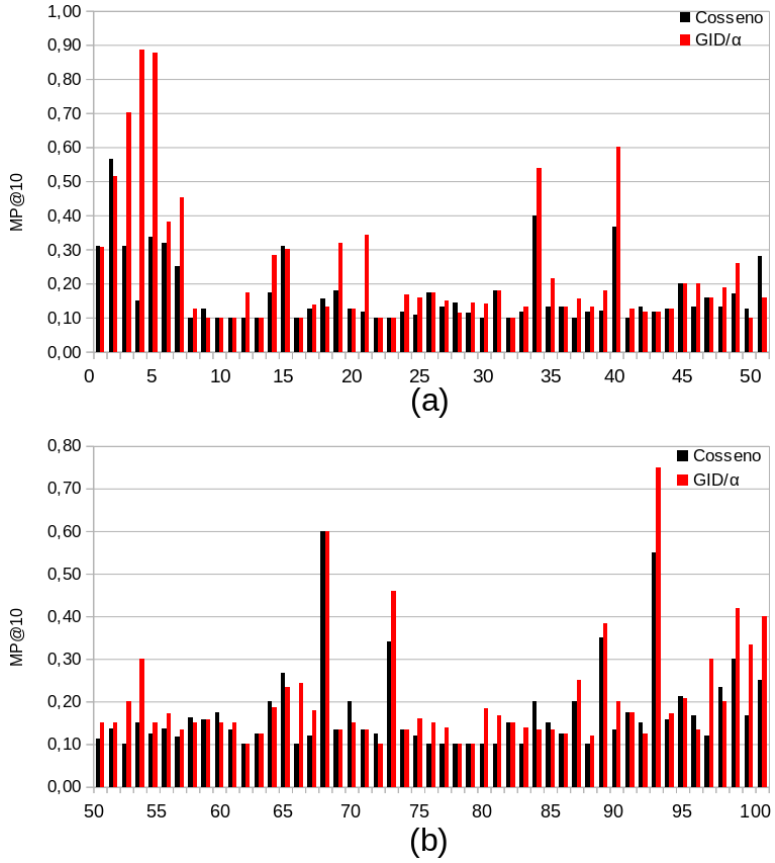

Figura 38 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Caltech101 com as funções de similaridade Cosseno e GID utilizando a abordagem BSM, sendo divididas as 101 classes em duas partes, da classe 1 até a classe 50 (a) e da classe 51 até 101 (b).

Banco Caltech101

Neste tópico são apresentados os resultados dos experimentos utilizando a base de dados Caltech101. Aplicou-se três níveis (0 até 2) das pirâmides espaciais com o *codebook* de tamanho 600, gerando histogramas que pertence ao conjunto *Simplex* (como mostrado na Eq. (50)), de tamanho igual a 12.000. As funções utilizadas foram: Euclidiana, Cosseno e a divergência KL para análise da similaridade. Os resultados apresentados na Tabela 6 são as médias dos resultados obtidos utilizando todas as 9.144 imagens do banco Caltech101 como consulta.

Neste experimento, os histogramas gerados podem apresentar frequências zero e, como os vetores estão normalizados em $[0, 1]$, foi aplicado o tratamento KL/ε com $\varepsilon = 0,00001$ (subseção 5.1) de acordo com teste empíricos. A Tabela 6 apresenta os resultados em termos de MP@10, MP@20, MP@30, nDCG@10, nDCG@20, nDCG@30, MAP e MAP@200 para o método BoVW-SPM. A porcentagem positiva significa a melhora da KL/ε comparada com o resultado obtido pela Cosseno que foi superior por aquele obtido usando a distância Euclidiana usada para medir a similaridade entre os histogramas.

Analisando os resultados ilustrados na Tabela 6, observa-se que o maior ganho foi

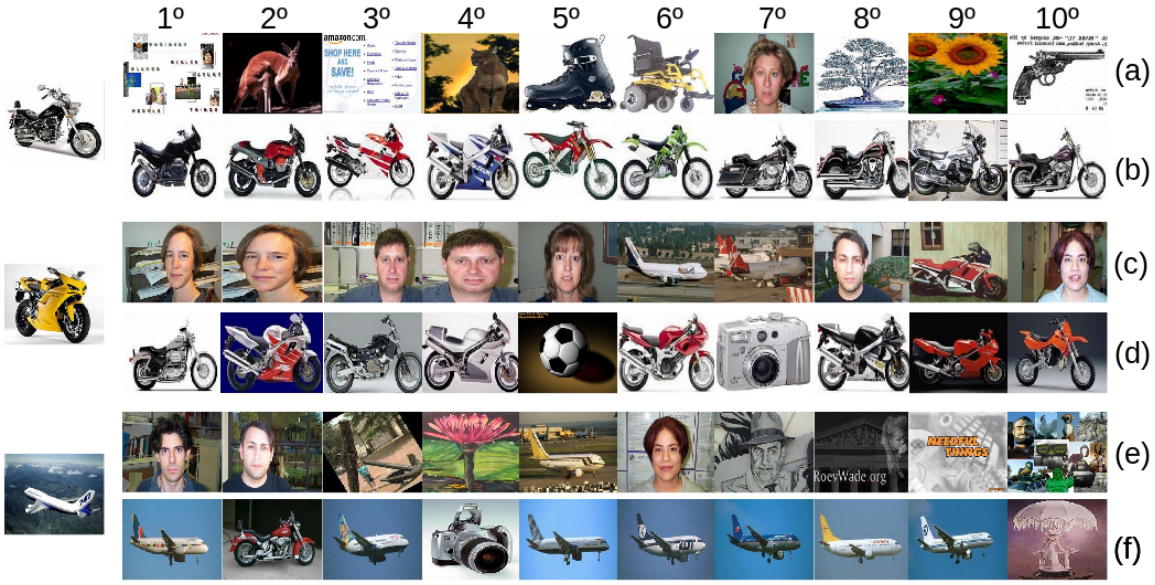

Figura 39 – Resposta da recuperação utilizando três imagens (diferentes) de consultas que não pertencem ao banco. Para cada consulta há duas listas ranqueadas das top 10 imagens da recuperação, a primeira linha mostra a recuperação utilizando a medida Cosseno (a; c; e) e a segunda linha a GID (b; d; f).

Tabela 6 – Resultados obtidos no banco Caltech101 com as funções Cosseno, Euclidiana e a KL/ε utilizando a abordagem BoVW-SPM.

Abordagem	BoVW-SPM		
Função de Similaridade	Euclidiana	Cosseno	KL/ε
MP@10	0,4172	0,4439	0,5847 +31,72%
MP@20	0,3481	0,3738	0,5149 +37,75%
MP@30	0,3148	0,3398	0,4766 +40,25%
nDCG@10	0,4913	0,5166	0,6419 +24,25%
nDCG@20	0,4220	0,4472	0,5779 +29,23%
nDCG@30	0,3846	0,4094	0,5396 +31,80%
MAP@200	0,5051	0,5213	0,5854 +12,30%
MAP	0,1305	0,1623	0,2917 +79,73%

de 79% considerando a caracterização BoVW-SPM quando utilizou a KL/ε em vez da medida Cosseno, e chegando ao ganho mínimo de 12% no MAP@200. Observando os resultados de forma geral, podemos perceber que a KL/ε foi superior em desempenho do que a medida Cosseno e a Euclidiana como função de similaridade em todos os métodos de avaliação.

Também foram analisadas a precisão e a revocação média da Tabela 6, os gráficos estão ilustrados na Figura 40. Os resultados obtidos com o uso função KL/ε apresentam resultados superiores se comparadas com o uso das medidas Cosseno e Euclidiana em todos os níveis de revocação. Mostrando novamente a qualidade dos resultados obtidos com a divergência de Bregman.

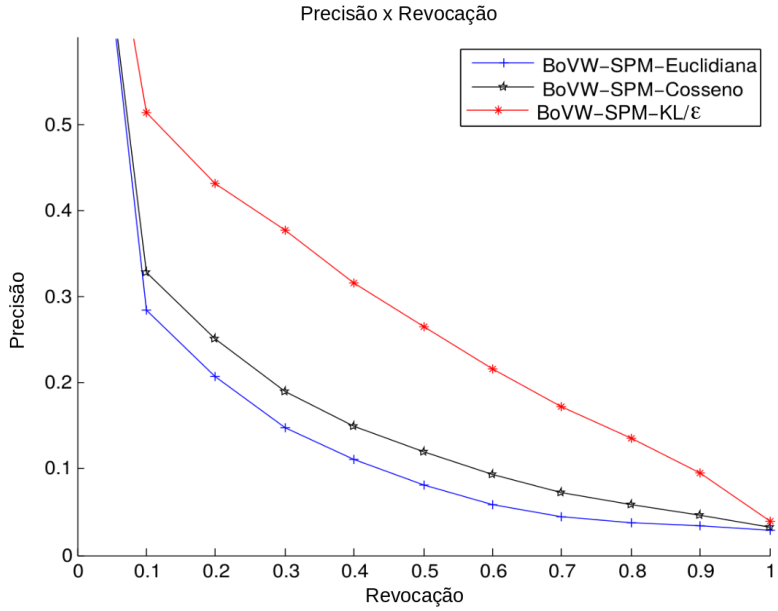

Figura 40 – Precisão x revocação média aplicando a abordagem BoVW-SPM com as medidas KL/ε, Cosseno e Euclidiana para o cálculo de similaridade no banco Caltech101.

A Figura 41 apresenta os resultados médio por classe das funções Cosseno e KL/ε da Tabela 6, utilizando a medida de avaliação MP@10. Na Figura 41 nota-se que o menor resultado obtido com a medida KL/ε foi da classe 9 (MP@10=0,1261) enquanto que o maior resultado foi da classe 7 (MP@10 = 0,9868). A classe 9 é chamada de *ant* (formiga) e tem um total de 42 imagens, enquanto que a classe 7 *airplanes* (aviões) contém um total de 800. As Figuras 42 e 43 mostram exemplos de algumas imagens da classe 7 e 9, respectivamente. Acredita-se que a pequena quantidade de imagens da classe 9 tenha influenciado pouco na geração do *codebook*, consequentemente os histogramas gerados para representar as imagens desta classe não foram representativos, além disso, as imagens desta classe contém formigas com diferentes perspectivas e padrões, sendo mais um motivo do resultado insatisfatório. Já a classe 7 contém um maior número de imagens, consequentemente aconteceu o oposto da classe 9, o maior número de imagens, de forma indireta, influenciou no dicionário de palavra visual, fazendo com que o histograma seja mais representativo, e também, as imagens desta classe têm um padrão no posicionamento do objeto (visão lateral do avião), sendo um provável motivo da qualidade dos resultados obtidos.

Avaliando o resultado médio por classe de avaliação MP@10, com o uso da similaridade Cosseno apresentada na Figura 41, o maior resultado obtido foi da classe 5 (MP@10=0,9159) enquanto que a menor avaliação foi da classe 9 (MP@10=0,1071). A classe 5 é nomeada de *motorbikes* (motos), sendo composta por 798 imagens de motos, a Figura 44 apresenta alguns exemplos desta classe. Acredita-se que a grande quantidade de imagens da classe 5 tenha influenciado na geração do dicionário de palavras, desta forma,

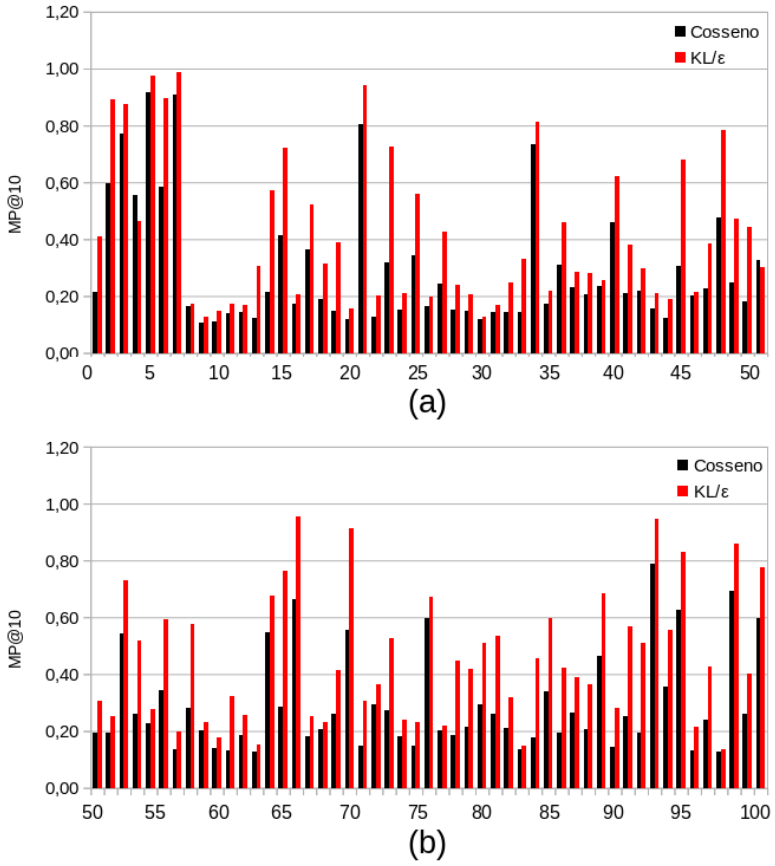

Figura 41 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Caltech101 com as funções de similaridade Cosseno e KL/ε utilizando a abordagem BoVW-SPM. As 101 classes estão divididas em duas partes, da classe 1 até a classe 50 (a) e da classe 51 até 101 (b).

Figura 42 – Exemplo de algumas imagens contidas na classe 7 (aviões).

os histogramas gerados das imagens desta classe foram melhores representadas do que a classe 9. Observe que a classe 9 obteve resultados insatisfatórios em ambas as medidas (Cosseno e KL/ε).

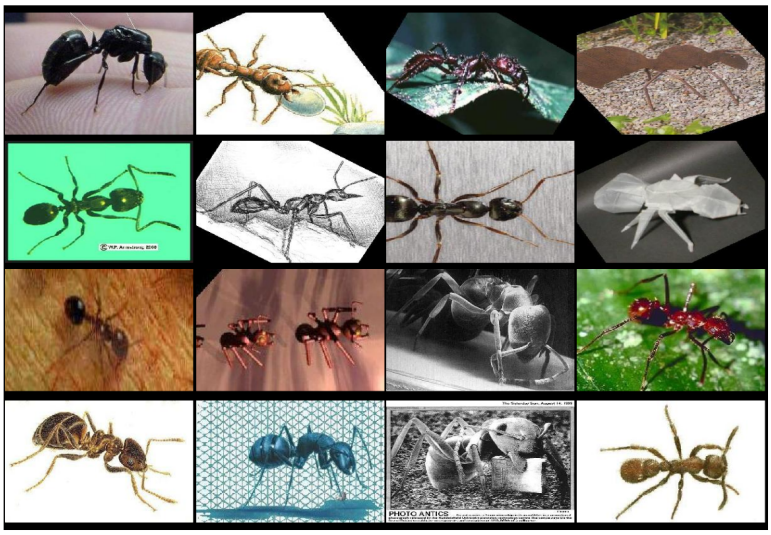

Figura 43 – Exemplo de algumas imagens contidas na classe 9 (formigas).

Figura 44 – Exemplo de algumas imagens contidas na classe 5 (motos).

Analisando mais detalhadamente a Figura 41, observa-se que apenas duas classes (4 e 51) utilizando a medida Cosseno como função de similaridade foram superiores do que a KL/ε. A classe 4 é chamada de *Leorpards* (Leopardo) e contém um total de 200 imagens, enquanto que a classe 51 é nomeada de *Hedgehog* (Ouriço) e contém um total de 54 imagens. As Figuras 45 e 46 apresentam algumas imagens da classe 4 e 51, respectivamente. O resultado do MP@10 da classe 4 e 51 utilizando a função Cosseno foram 0,5535 e 0,3259 respectivamente, já com o uso da KL/ε obteve-se as avaliações 0,4645 e 0,3018 para as classes 4 e 51 na devida ordem. No entanto, 98% das classes do banco, de um total de 101, o uso da divergência (KL) como medida de similaridade mostrou-se superior ao uso da medida Cosseno.

Por fim, a última tabela deste tópico (Tabela 7) reflete os resultados obtidos com 30% de imagens de cada classe como consulta, utilizando a mesma abordagem (BoVW-SPM), funções de similaridade e parâmetros descritos no início deste tópico. Entretanto,

Figura 45 – Exemplo de algumas imagens contidas na classe 4 (Leopardo).

Figura 46 – Exemplo de algumas imagens contidas na classe 51 (Ouriço).

o dicionário de palavras visuais foi construído apenas com as características extraídas das imagens do banco de dados, desconsiderando as informações das imagens de consulta. A Tabela 7 apresenta os resultados em termos de MP@10, MP@20, MP@30, nDCG@10, nDCG@20, nDCG@30 e MAP, para a abordagem BoVW-SPM.

Com base na analise da Tabela 7, observamos que em todos os métodos da avaliação a divergência KL/ε obteve melhores resultados do que a distância Euclidiana e a similaridade Cosseno. A divergência conseguiu um ganho máximo 83,25% com avaliação MAP e um ganho mínimo de 23,95% com avaliação nDCG@10, quando comparada com a similaridade Cosseno. Observe também que os resultados apresentados nas Tabelas 6 e 7 são próximos, utilizando as medidas de avaliação MP, nDCG e MAP.

Os gráficos de precisão e a revocação média da Tabela 7 estão ilustrados na Figura 47. Os resultados da similaridade Cosseno e a distância Euclidiana são inferiores que a divergência KL/ε em todos os níveis de revocação. Também destacamos que as Figuras

Tabela 7 – Resultados obtidos no banco Caltech101 com as funções Cosseno, Euclidiana e a KL/ε utilizando a abordagem BoVW-SPM. Como consulta foi utilizado 30% de imagens de cada classe. O dicionário de palavras visuais foi construído utilizando apenas as imagens do banco de dados, descartando as informações das imagens de consulta.

Abordagem	BoVW-SPM		
Função de Similaridade	Euclidiana	Cosseno	KL/ε
MP@10	0,4176	**0,4446**	0,5825 +31,02%
MP@20	0,3489	**0,3722**	0,5171 +38,93%
MP@30	0,3164	**0,3388**	0,4791 +41,41%
nDCG@10	0,4910	**0,5160**	0,6396 +23,95%
nDCG@20	0,4222	**0,4455**	0,5785 +29,85%
nDCG@30	0,3854	**0,4081**	0,5407 +32,49%
MAP	0,1311	**0,1594**	0,2921 +83,25%

40 e 47 conseguiram resultados bastante similares, mesmo com o dicionário de palavras visuais sendo construído de modos diferentes.

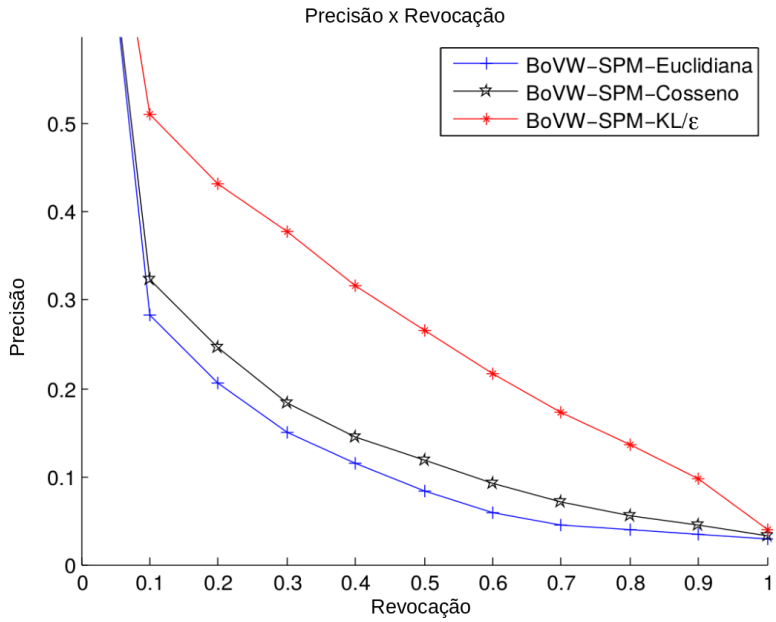

Figura 47 – Precisão x revocação média aplicando a abordagem BoVW-SPM com as medidas KL/ε, Cosseno e Euclidiana para o cálculo de similaridade no banco Caltech101. Como consulta foi utilizado 30% de imagens de cada classe. O dicionário de palavras visuais foi construído utilizando apenas as imagens do banco de dados, descartando as informações das imagens de consulta.

Também foram analisadas os resultados médio por classe das funções Cosseno e KL/ε da Tabela 7, usando a medida de avaliação MP@10, os resultados estão apresentados na Figura 48. Analisando a Figura 48, observa-se que a divergência de Bregman conseguiu resultados superiores em 96 (95,05%) classes do total do banco. Já a Cosseno obteve resultados superiores apenas em 3 classes, que corresponde 2,97% do banco. Ambas

medidas de similaridade (KL/ε e Cosseno) conseguiram resultados iguais em 2 classes. De modo geral, o uso da KL/ε como função de similaridade mostrou-se superior que a medida Cosseno, tendo resultados melhores na maioria das classes do banco Caltech101, mostrando mais uma vez que a divergência KL pode ser uma boa opção como função de similaridade na etapa de recuperação.

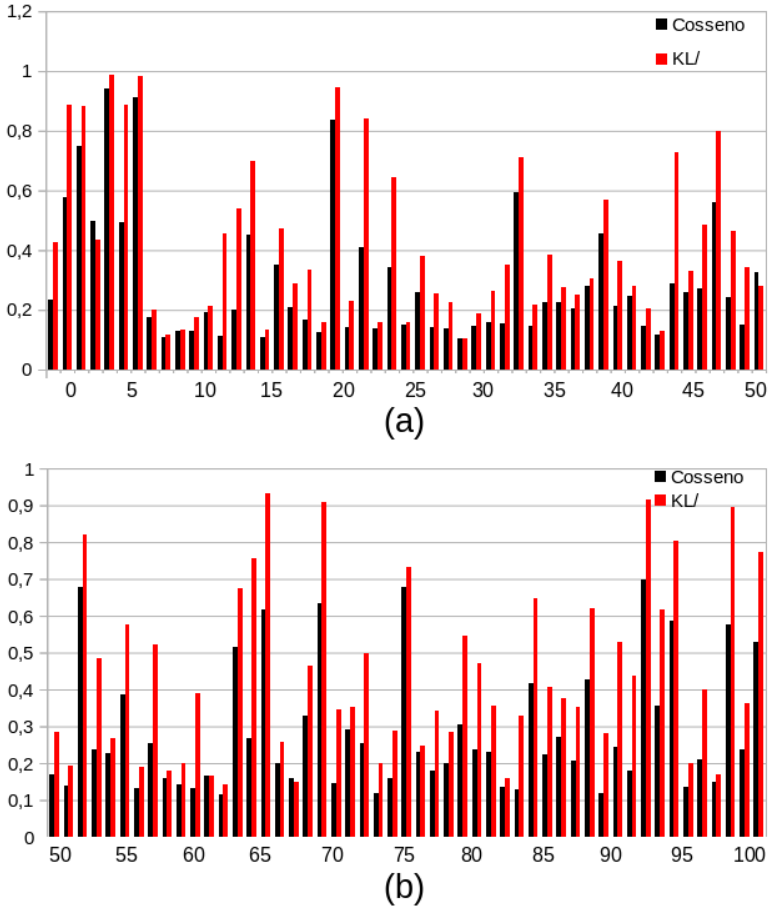

Figura 48 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@10 obtidos no banco Caltech101 com as funções de similaridade Cosseno e KL/ε utilizando a abordagem BoVW-SPM. As 101 classes estão divididas em duas partes, da classe 1 até a classe 50 (a) e da classe 51 até 101 (b).

Banco UK-Bench

Os mesmos experimentos (descrito no tópico Banco Caltech101 – Experimento II) com os mesmos parâmetros foram conduzidos no banco UK-Bench. A Tabela 8 apresenta os resultados em termos de MP@1, MP@2, MP@3 e MP@4[6] para avaliar o desempenho das funções de similaridade Euclidiana, Cosseno e a divergência KL/ε, com $\varepsilon = 0,00001$. A porcentagem apresentada está relacionada ao ganho da divergência KL comparada com

[6] É utilizado até MP@4 devido a característica do banco UK-bench contém apenas quatro imagens por classe como descrito na subseção 7.1.

a distância Euclidiana, que foi superior aquela obtida com medida Cosseno usada para medir a similaridade.

Tabela 8 – Resultados obtidos no banco UK-bench com as funções Cosseno, Euclidiana e a KL/ε utilizando a abordagem BoVW-SPM.

Abordagem	BoVW-SPM		
Função de Similaridade	Euclidiana	Cosseno	KL/ε
MP@1	**1,0000**	1,0000	1,0000 +00,00%
MP@2	**0,6575**	0,6525	0,8325 +26,62%
MP@3	**0,4925**	0,4875	0,7075 +43,65%
MP@4	**0,3900**	0,3850	0,6025 +54,49%

Analisando a Tabela 8, observamos que a utilização da divergência KL obteve um ganho de até 54,49%, comparado com as outras medidas analisadas evidenciando mais uma vez a superioridade das divergências em relação às distâncias tradicionais como, por exemplo, a Euclidiana e a Cosseno. Os resultados do MP@1 para todas as funções de similaridade foram iguais a 1 porque a imagem de consulta (que também está contida no banco) sempre é retornada como ela própria sendo a mais similar.

Para auxiliar avaliação das medidas de similaridade no Banco UK-bench, também foram analisadas a precisão e a revocação média dos resultados da Tabela 8 e os gráficos estão ilustrados na Figura 49. Observe que em todos os níveis de revocação a divergência KL/ε obteve precisão igual (níveis até 20% de revocação) ou superior (maior do que 20% de revocação) do que as medidas tracionais (Cosseno e Euclidiana). Estes resultados mostram que a divergência KL pode ser uma boa escolha como função de similaridade, no contexto de CBIR.

A Figura 50 apresenta avaliação MP@4 por classe de todo o banco UK-bench, utilizando como função de similaridade a distância Euclidiana (Figura 50a) e a divergência KL/ε (Figura (50b)). Com base na análise do gráfico, apenas 3,76% das classes, que corresponde a 96 classes de um total de 2.550, utilizando a distância Euclidiana como medida de similaridade foram superiores do que a divergência KL/ε. Enquanto que 2.005 classes (que corresponde a 78,63%) do banco a divergência KL/ε obteve resultados melhores do que a distância Euclidiana. As 449 classes restantes, que representa 17,61% do total do banco, a distância Euclidiana e a divergência KL/ε conseguiram resultados iguais. De forma geral, a divergência KL alcançou avaliações melhores do que a distância Euclidiana nas diferentes classes do banco de dados UK-bench.

Para uma análise mais detalhadas da avaliação MP@4 por classe do banco UK-bench com as medidas de similaridade Euclidiana e KL/ε, foram selecionadas duas amostras da Figura 50, que são as Figuras 51 e 52.

A Figura 51 apresenta a performance da avaliação MP@4 das classes de 51 até 100 no banco UK-bench. A KL/ε foi superior do que a Euclidiana em 92% das classes (51 até

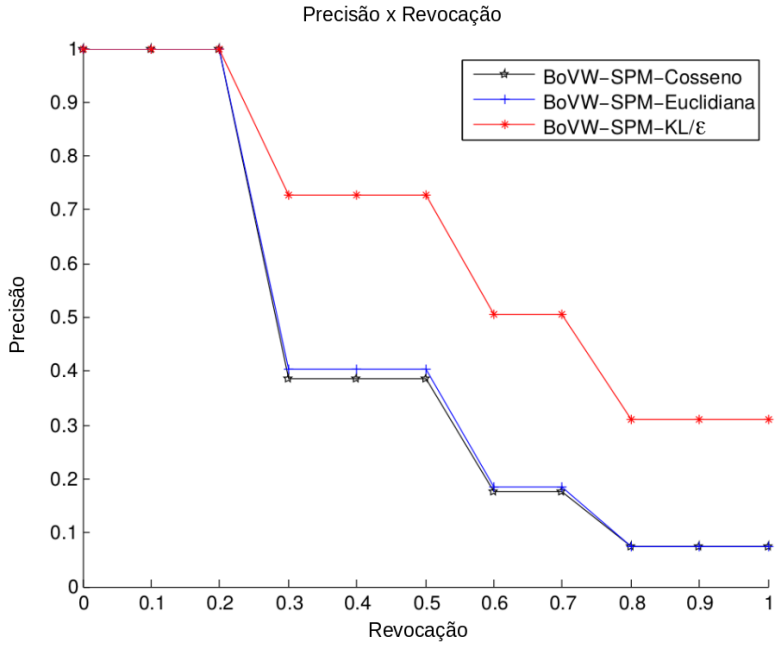

Figura 49 – Precisão x revocação média aplicando a abordagem BoVW-SPM com as medidas KL/ε, Cosseno e Euclidiana para o cálculo de similaridade no banco UK-bench.

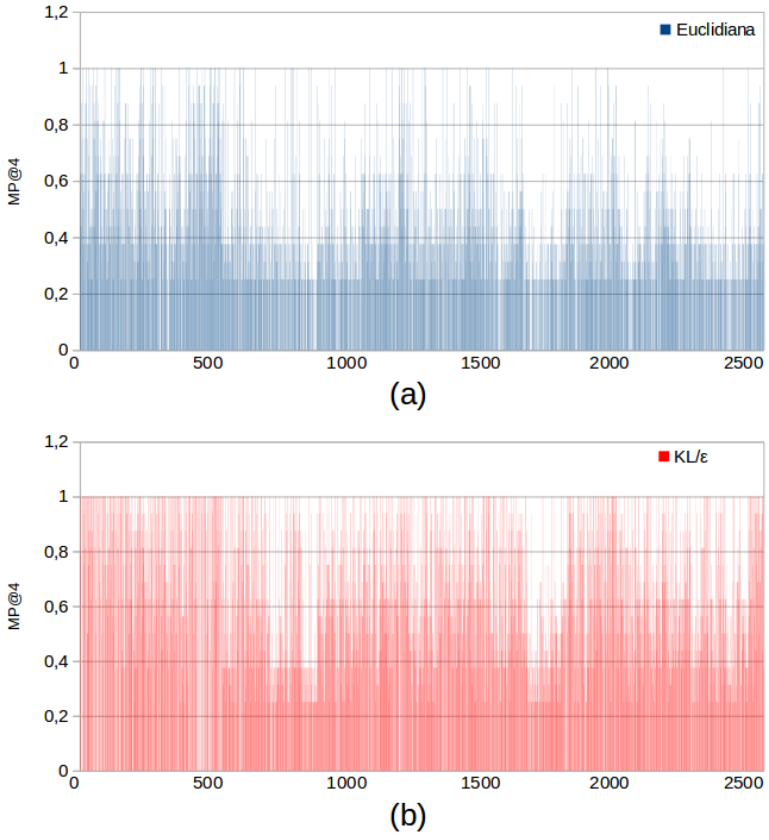

Figura 50 – Gráfico de barras da performance por classe (eixo x) da avaliação MP@4 obtidos no banco UK-bench com as funções de similaridade Euclidiana (a) e KL/ε (b) utilizando a abordagem BoVW-SPM.

100) e conseguiram resultados equivalentes 8%, ou seja, a Euclidiana nesse intervalo de classes (51 até 100) do UK-bench não conseguiu nenhuma avaliação superior que a KL/ε.

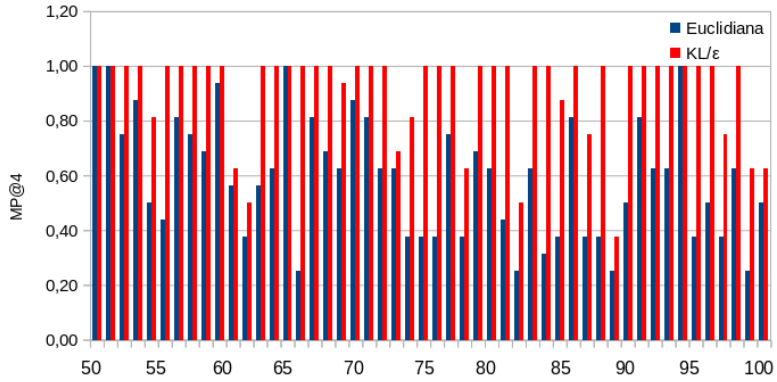

Figura 51 – Gráfico de barras da performance das classes de 51 até 100 (eixo x) da avaliação MP@4 obtidos no banco UK-bench com as funções de similaridade Euclidiana e KL/ε utilizando a abordagem BoVW-SPM.

Por fim, a Figura 52 ilustra os resultados da avaliação MP@4 nas classes de 701 até 750 no Banco UK-bench. A divergência KL/ε conseguiu resultados superiores do que a distância Euclidiana em 52% (26 classes do total de 50), enquanto que o uso da Euclidiana alcançou resultados melhores em 18% do total das classes (no intervalo de 701 até 750), ambas medidas conseguiram resultados iguais em 30% das classes analisadas.

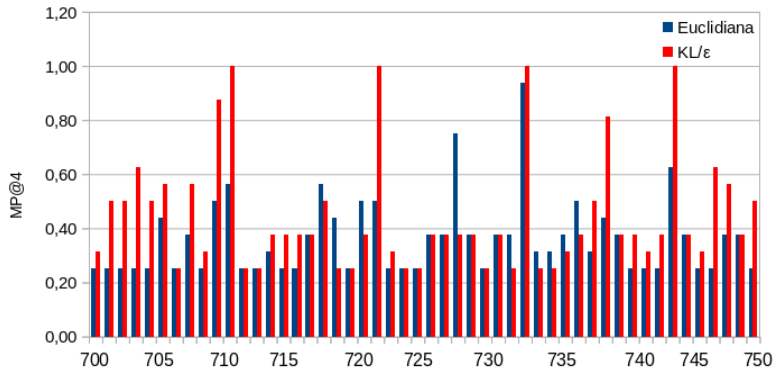

Figura 52 – Gráfico de barras da performance das classes 701 até 750 (eixo x) da avaliação MP@4 obtidos no banco UK-bench com as funções de similaridade Euclidiana e KL/ε utilizando a abordagem BoVW-SPM.

Banco Holiday

Neste tópico são apresentados os experimentos no banco de dados Holiday. Os parâmetros utilizando são o mesmo descrito no tópico Banco Caltech101 (Experimento II). As funções utilizadas como medida de similaridade foram a Euclidiana, Cosseno e a divergência KL/ε (com $\varepsilon = 0{,}00001$).

As imagens de consultas utilizadas foram todas as imagens que formam o banco, ou seja, as 1.491 imagens. A Tabela 9 apresenta os resultados obtidos com a avaliação MAP,

para avaliar o desempenho das medidas Euclidiana, Cosseno e KL/ε. Foi utilizada apenas avaliação MAP devido o número de imagens por classe variar de 2 até 13 imagens. A porcentagem positiva é o ganho da divergência KL/ε em relação a medida Cosseno, pois a Cosseno foi superior que a distância Euclidiana. Observando os resultados da Tabela 9, vemos que a KL/ε conseguiu ganhos de 36,46%, comparada com similaridade Cosseno, apresentando resultados com qualidade no banco Holiday, quando comparados com as medidas tracionais (Euclidiana e Cosseno).

Tabela 9 – Resultados obtidos no banco Holiday com as funções Cosseno, Euclidiana e a KL/ε utilizando a abordagem BoVW-SPM.

Abordagem	BoVW-SPM		
Função de Similaridade	Euclidiana	Cosseno	KL/ε
MAP	0,5142	**0,5242**	0,7152 +36,46%

Também analisamos a precisão e a revocação média da Tabela 9, os gráficos estão apresentados na Figura 53. Veja que em todos os níveis de revocação a divergência KL/ε foi superior do que a distância Euclidiana e a medida Cosseno. De forma geral, os experimentos realizado no banco Holiday, a divergência KL/ε conseguiu resultados superiores do que as medidas Euclidiana e Cosseno, quando utilizada para medir a similaridade entre histogramas.

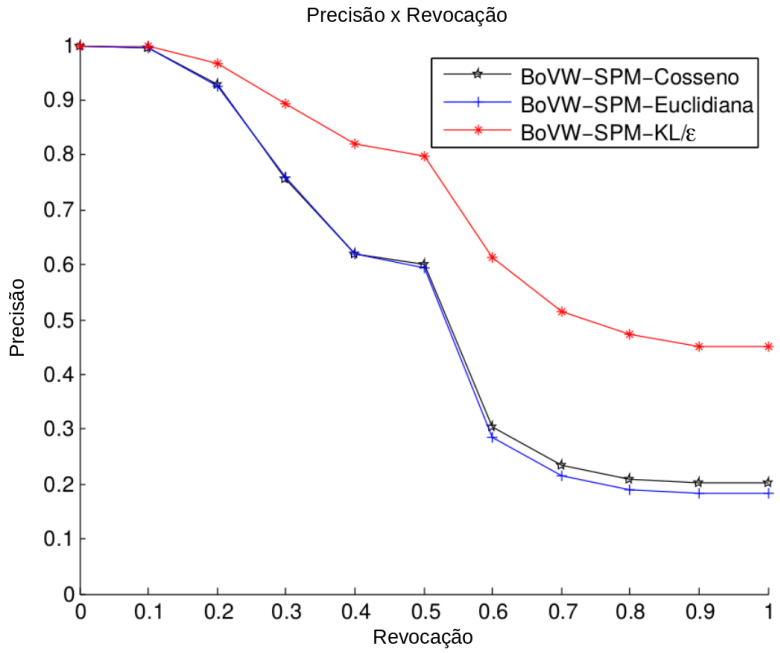

Figura 53 – Precisão x revocação média aplicando a abordagem BoVW-SPM com as medidas Cosseno, Euclidiana e a KL/ε para o cálculo de similaridade no banco Holiday.

7.3.2 Grupo 2 – experimentos III

As divergências GID e KL utilizam funções logarítmicas que só são definidas para dados positivos. E, como, em geral, a caracterização contém dados nulos, exige-se uma análise matemática das expressões GID e KL (mostradas na Tabela 2).

Este experimento aborda os resultados obtidos por quatro tratamentos, dentre eles, um usualmente utilizado na literatura (KL/TI – Eq. (51)) e três propostas criadas neste trabalho (KL/ε[7], KL/N – Seção 5.1 – e GID/D, GID/ε – Seção 5.1). Além disso, são apresentados os resultados obtidos com as medidas Cosseno e Euclidiana.

Para esses experimentos utilizou-se: o banco Caltech101 com a caracterização BoVW e BoVW-SPM, a mesmas usadas nos experimento I – grupo 1 (tópico 7.3.1.1) e experimentos II – grupo 1 (tópico 7.3.1.2), respectivamente. A caracterização BoVW apresenta vetores não pertencentes ao *Simplex*, o que justifica o uso da GID. Para os experimentos foram utilizados como consulta 10% das imagens de cada classe, para análise da KL/ε, KL/N, GID/D, GID/ε bem como para a KL/TI.

Para cada tratamento realizado nas funções de Bregman (KL e GID), são possíveis configurar os parâmetros (γ, ε, β, α – veja a Seção 5.1) para adaptar os tratamentos de acordo com as características dos dados. Os resultados obtidos na Tabela 10 foram utilizados os seguintes parâmetros: $\varepsilon = 0,25$ e $\alpha = 1$. Enquanto que os resultados apresentados na Tabela 11 foram usados os parâmetros: $\gamma = 10$, $\varepsilon = 0,00001$, $\beta = 3,75$, $\alpha = 1/d$, onde d é o tamanho do histograma, neste caso $d = 12.000$. Observe que os mesmos parâmetros da função KL/ε foram utilizados para as Tabelas 6 e 11, porém, uma utiliza todas imagens do banco como consultas (Tabela 6) e a outra não (Tabela 11).

Os resultados obtidos com a abordagem BoVW utilizando-se as métricas de avaliações MAP; MP@10, MP@20 e MP@30; e o nDCG para as primeiras 10 (nDCG@10), 20 (nDCG@20) e 100 (nDCG@100) respostas da lista ranqueada (MANNING; RAGHAVAN; SCHÜTZE, 2008) e os tratamentos acima descritos são apresentados na Tabela 10. A porcentagem positiva informa a melhora dos tratamentos para a divergência em relação à medida Cosseno. Vale ressaltar que não foram mostrados os resultados da distância Euclidiana por serem inferiores à Cosseno, como também, não foram apresentados os resultados da divergência KL, devido os dados não estarem no domínio *Simplex*.

Analisando os resultados mostrados na Tabela 10, nota-se que os tratamentos (GID/D e GID/ε) para os dados, conseguiu um ganho máximo de 51% para a avaliação MP@30 comparada com a medida Cosseno, e um ganho mínimo de 24% na avaliação nDCG@10. De forma geral a GID/D e GID/ε apresentou bons resultados quando comparados com a medida Cosseno, mostrando que estes tratamentos podem ser uma boa escolha para o uso da função GID.

O impacto do parâmetro ε na função GID/ε comparado com a medida Cosseno com a

[7] KL/ε e o GID/ε são equivalente quando os dados estão normalizados, e estão sendo considerados como um único tratamento.

Tabela 10 – 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%

abordagem BoVW está apresentado na Figura 54. Nota-se que, dependendo ε, a avaliação do MP@10 pode ser significativa com $0,01 \leq \varepsilon \leq 0,25$ e muito inferior à função Cosseno, como por exemplo $\varepsilon = 10$. Desta forma, com uma boa escolha para o valor de ε pode-se chegar a um ganho de 33% sobre a medida Cosseno.

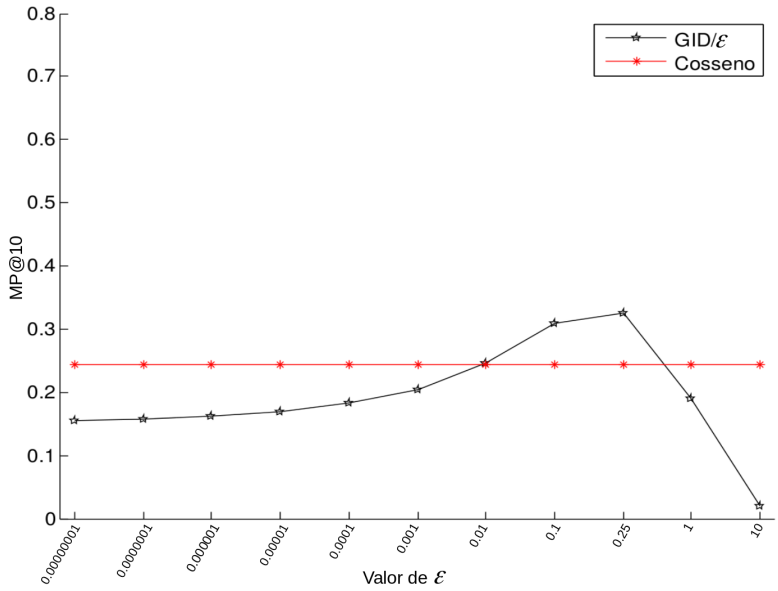

Figura 54 – O impacto do parâmetro ε na função GID/ε com a medida de avaliação MP@10, no banco de dados Caltech101 com a abordagem BoVW.

Enquanto que a Figura 55 apresenta o impacto do parâmetro α na função GID/D comparando com a medida Cosseno com a aproximação BoVW, nota-se que também é necessária uma boa escolha do α para obter ganhos de até 51% em relação a Cosseno, sendo que, a partir do $\alpha \geq 1.000$, o gráfico torna-se estável com a avaliação MP@10 e, ao mesmo tempo, superior à função Cosseno.

A Tabela 11 apresenta os resultados em termos de MAP, MP@10, MP@20, MP@30; nDCG@10, nDCG@20, nDCG@30 para a abordagem BoVW-SPM com os tratamentos KL/ε, KL/N, GID/D e KL/TI. A porcentagem positiva e negativa indica a melhora ou piora respectivamente dos tratamentos para as divergências em relação à medida Cosseno.

Analisando os resultados mostrados na Tabela 11, observa-se que os tratamentos propostos neste trabalho comparados com a função Cosseno obtiveram um ganho mínimo

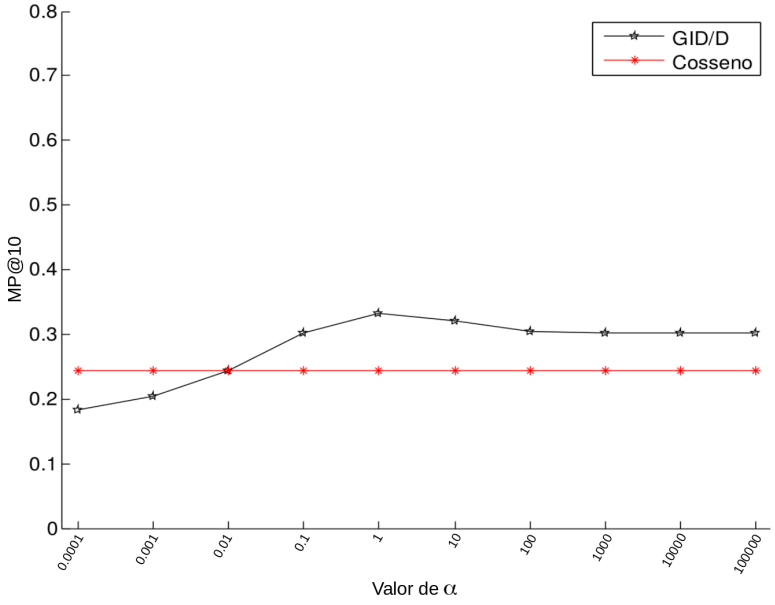

Figura 55 – O impacto do parâmetro α na função GID/D com a medida de avaliação MP@10, no banco de dados Caltech101 com a abordagem BoVW.

Tabela 11 – 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%

de 23% entre os tratamentos KL/ε, KL/N e GID/D, e um ganho máximo de 86% utilizando a função KL/ε na avaliação MAP. Em média, todos os tratamentos propostos (GID/D, KL/ε e KL/N) neste trabalho, apresentaram bons resultados, excedendo os resultados apresentados pelas distâncias Euclidiana e Cosseno, mostrando que o uso das divergências com os tratamentos pode se tornar vantajoso, desde que os parâmetros estejam configurados adequadamente.

A Figura 56 apresenta o impacto do parâmetro ε na função KL/ε comparado com a função Cosseno e a KL/TI com o $\gamma = 10$ fixo, nota-se que o parâmetro ε influencia na avaliação do resultado do MP@10, sendo que $0,0000001 \leq \varepsilon \leq 0,0001$ a medida de avaliação mantém os resultados superiores (possibilitando ganhos de até 36%) comparados com a medida Cosseno. Entretanto, a escolha do ε pode interferir no desempenho da função como por exemplo $\varepsilon \geq 0,001$, que apresentam resultados insatisfatórios para a avaliação MP@10.

Enquanto que a Figura 57 mostra a avaliação do MP@10 com a variação do valor de α na função GID/D comparando com a KL/TI e a medida Cosseno, veja-se que

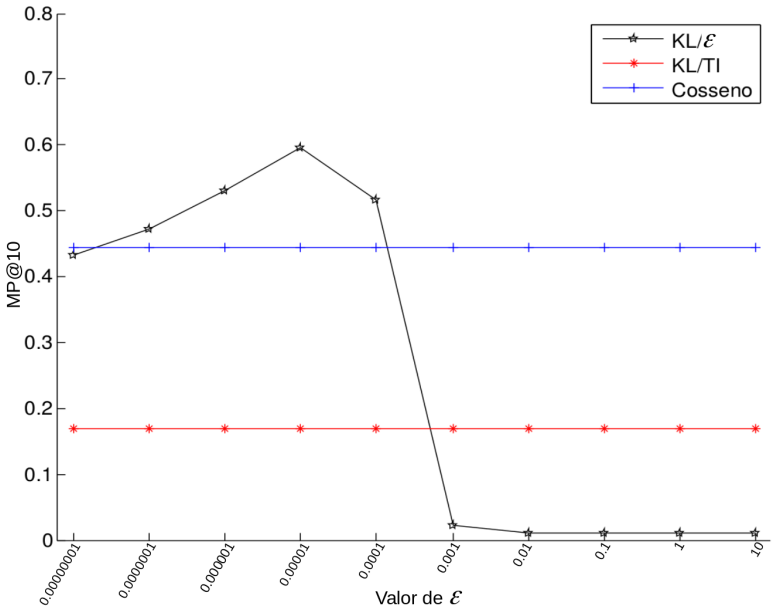

Figura 56 – O impacto do parâmetro ε na função KL/ε com a medida de avaliação MP@10, no banco de dados Caltech101.

$0,0000001 \leq \alpha \leq 0,01$ mostra-se superior a função Cosseno, e que, dependendo do valor para α, o desempenho da GID/D não é favorável, por exemplo $\alpha = 10$.

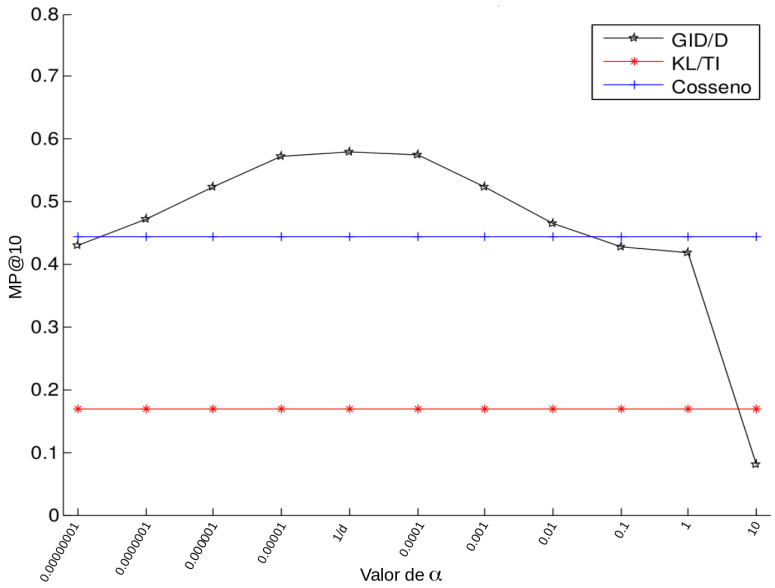

Figura 57 – O impacto do parâmetro α na função GID/D com a medida de avaliação MP@10, no banco de dados Caltech101.

Ao avaliar os resultados obtidos na Tabela 11 com o KL/TI (tratamento apresentando em (COVER; THOMAS, 1991)), foram investigados de forma detalhada os motivos dos resultados insatisfatórios mostrados. Primeiramente, analisaram-se os histogramas que representam as imagens do banco, gerados pela aproximação BoVW-SPM, e notou-se que os vetores são compostos em média de 70% de suas coordenadas iguais a zeros, fazendo com que a KL/TI com o seu tratamento ($x_j \neq 0$ e $y_j = 0$) venha a interferir nos resultados

e tornando-os incertos. Considerando quanto maior o resultado da função $d_{KL/TI}(\mathbf{x}, \mathbf{y})$, menos similar é a imagem de consulta com a imagem da base, o resultado da $d_{KL/TI}(\mathbf{x}, \mathbf{y})$ será extremamente alto devido às coordenadas $\mathbf{y}$ conterem uma grande quantidades de 0's e, consequentemente, as parcelas $x_i log_2(\frac{x_i}{y_i}) = \gamma$, onde γ é um valor muito alto para que seja computável e, ao somarem-se todas as parcelas da $d_{KL/TI}(\mathbf{x}, \mathbf{y})$ resultará em um valor extremamente alto. Com isso, na recuperação da imagem (representada por $\mathbf{y}$), normalmente ficaria localizada nas últimas posições da lista de imagens ranqueadas como menos similar. Caso contrário (o vetor $\mathbf{y}$ tem poucos 0's em suas coordenadas), a $d_{KL/TI}(\mathbf{x}, \mathbf{y})$ resultará em um valor muito menor, fazendo com que a imagem se localize nas primeiras posições da lista ranqueada. A Figura 58 mostra um exemplo de recuperação de três imagens de consultas diferentes contida no banco Caltech101, onde são retornadas as 10 primeiras imagens mais semelhantes comparadas com a imagem de consulta, a primeira e a segunda linha representam a recuperação KL/N e KL/TI respectivamente, e pode ser notado que em diferentes consultas, as mesmas imagens sempre são retornadas, supostamente, as que são representadas por histogramas que contém um menor número de 0's em suas coordenadas. Mostrando assim a eficácia dos tratamentos descritos neste trabalho.

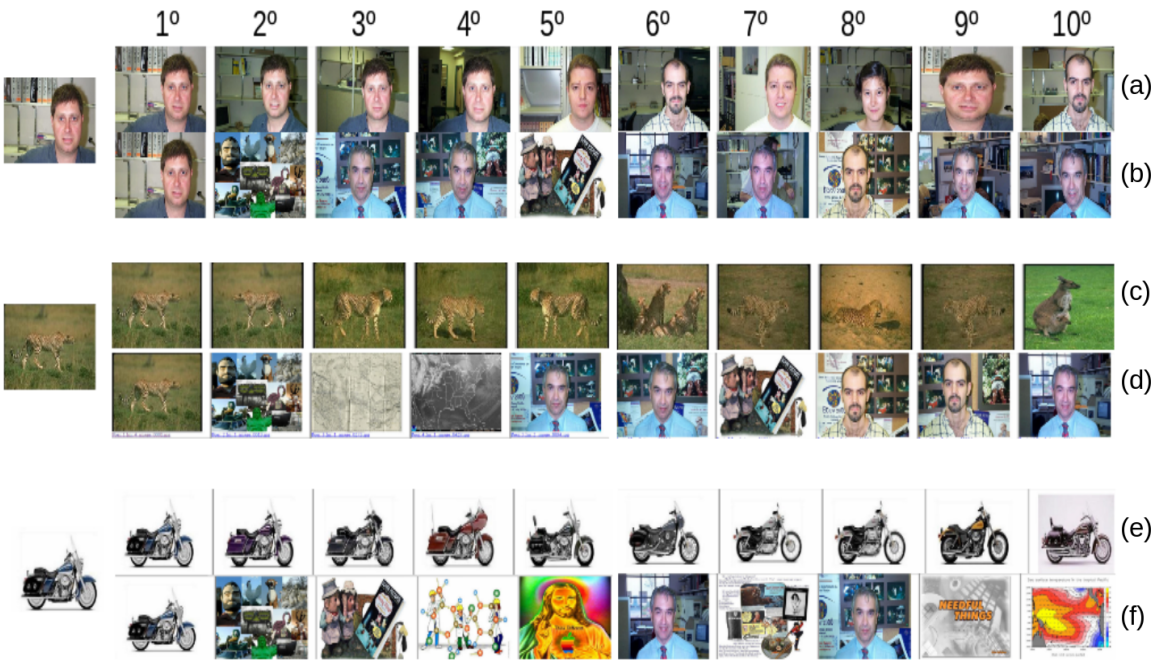

Figura 58 – Para cada consulta apresenta duas listas ranqueadas das top 10 imagens da recuperação, utilizando a divergência KL com dois tratamentos diferentes como medida de similaridade. Para as linhas (a; c; e) são as respostas da recuperação KL/N, enquanto que o uso da KL/TI são apresentados em (b; d; f).

7.3.3 Grupo 3 – experimentos IV

Esta seção tem como objetivo comparar os resultados obtidos nesse trabalho com dois trabalhos presentes na literatura, que são: o trabalho (XU et al., 2012) (utilizando o banco Caltech101) e (JÉGOU; CHUM, 2012) (com o banco Holiday). A seguir são descritos como foram conduzidos os experimentos, com intuito de compara-los, no banco Caltech101, posteriormente é explanado a condução dos experimentos no banco Holiday.

Banco Caltech101

Neste tópico são comparados os resultados obtidos no trabalho do (XU et al., 2012) com a proposta deste trabalho, utilizando o banco Caltech101. São apresentados os resultados obtidos por (XU et al., 2012) e nossos melhores resultados (utilizando a função KL/ε com a abordagem BoVW-SPM).

Em (XU et al., 2012) foi proposta uma divergência de Bregman para *manidold ranking* (MR). Os autores utilizaram como caracterização, um vetor de dimensão 297 considerando a característica *Grid Color Moment* (resultando um vetor de dimensão 81 para cada imagem); mapa de borda usando o detector de borda Canny (vetor de dimensão 37); As transformadas de *Gabor Wavelets* são aplicadas para imagens em escalas 64×64 com 5 níveis, 8 orientações e 3 momentos foram calculados, criando um vetor de dimensão 120. Também foi incluído na caracterização o *Local Binary Pattern* (OJALA; PIETIKÄINEN; HARWOOD, 1996) obtendo um vetor de dimensão 59. Com estas características e uma matriz de *kernel* otimizado sob a divergência de Bregman o modelo atingiu os resultados apresentados na Tabela 3 (página 79).

Para os experimentos foram utilizadas todas as imagens do banco como consulta (igualmente apresentando em (XU et al., 2012)), totalizando 9.144 imagens como consulta. O processo de criação dos histogramas de nosso trabalho e seus parâmetros são os mesmos descritos no experimento II (tópico Banco Caltech101), utilizando a abordagem BoVW-SPM. Mesclamos os melhores resultados da Tabela 3 do trabalho (XU et al., 2012) com os resultados obtidos neste trabalho (Tabela 6), resultando a Tabela 12. A porcentagem positiva significa o ganho da função KL/ε comparado com DMR_E.

Analisando a Tabela 12 observa-se que o método proposto é muito mais simples que o método de aprendizado apresentado por (XU et al., 2012), além disso, o uso da divergência KL/ε alcança resultados melhores em todos os métodos de avaliações apresentados, quando comparados com o DMR_E. Além disso, pode-se destacar ganhos mínimo de 41,30% e ganho máximo de 59,73% nos métodos MP@30 e MAP@200, respectivamente.

Banco Holiday

Neste tópico são apresentados os resultados obtidos neste trabalho utilizando o banco de dados Holiday e comparados com o trabalho de (JÉGOU; CHUM, 2012). No trabalho

Tabela 12 – Resultados obtidos no banco Caltech101 com algoritmo eficiente DMR_E ((XU et al., 2012)) e a KL/ε utilizando abordagem BoVW-SPM.

Abordagem	Método sem restrições	BoVW-SPM
Função de Similaridade	DMR_E	KL/ε
MP@10	0,3893	0,5847 +50,19%
MP@20	0,3572	0,5149 +44,15%
MP@30	0,3373	0,4766 +41,30%
nDCG@10	0,4077	0,6419 +57,44%
nDCG@20	0,3787	0,5779 +52,60%
nDCG@30	0,3600	0,5396 +49,89%
MAP@200	0,3665	0,5854 +59,73%

de (JÉGOU; CHUM, 2012) é proposto uma técnica para melhorar a qualidade da representação da abordagem BoVW. Para isto, é considerada o papel da *negative evidence*, que funciona da seguinte forma: dados dois histogramas gerados pelo BoVW, uma palavra visual que está faltando nestes dois vetores receber mais importância na medida de similaridade.

(JÉGOU; CHUM, 2012) criou o histograma BoVW da seguinte forma: em vez de utilizar apenas o descritor local SIFT como apresentado na seção 3.1, os autores combinaram o SIFT com o detector Hessian-Affine (MIKOLAJCZYK; SCHMID, 2005), para representar os pontos de interesse do conjunto de imagens; em seguida é construído o dicionário de palavras visuais utilizando o algoritmo de agrupamento *k-means* (com k=32.000); depois são criados os histogramas de ocorrências de palavras visuais e ponderados utilizando termos de frequência de documentos inversos (*inverse document frequency*) (JÉGOU; CHUM, 2012); e por fim os vetores são normalizados utilizando a normalização L2 (SIVIC; ZISSERMAN, 2003). A função Cosseno foi utilizada para medir a similaridade entre os histogramas, na etapa de recuperação. Chamaremos aqui a abordagem de (JÉGOU; CHUM, 2012) de $BoVW_J$.

Para efeito de comparação desta pesquisa com o trabalho de (JÉGOU; CHUM, 2012), utilizamos as mesmas imagens de consulta e a mesma medida de avaliação, mas com formas de caracterização diferentes. As imagens de consultas utilizadas foram a primeira imagem de cada classe do banco Holiday, ou seja, um total de 500 imagens. A medida de avaliação utilizada foi o MAP. Utilizamos a abordagem BoVW-SPM para gerar o histograma, com os mesmos parâmetros descritos no tópico Banco Caltech101 – Experimento II. As funções de similaridades utilizadas foram a Euclidiana, Cosseno e a divergência KL/ε, com ε=0,00001.

A Tabela 13 mostra os resultados alcançados com a avaliação MAP, usando a similaridade Cosseno com abordagem $BoVW_J$ – como apresentando em (JÉGOU; CHUM, 2012), e as funções Euclidiana, Cosseno e KL/ε utilizando a abordagem BoVW-SPM. A porcentagem positiva e negativo são o ganho e perda respectivamente dos resultados

conseguidos com a abordagem BoVW-SPM em relação a BoVW$_J$.

Tabela 13 – Resultados obtidos no banco Holiday com as funções Cosseno, Euclidiana e a KL/ε utilizando as abordagens BoVW$_J$ e BoVW-SPM.

Abordagem	BoVW$_J$	BoVW-SPM		
Função de Similaridade	Cosseno	Euclidiana	Cosseno	KL/ε
MAP	**0,6000**	0,5655 -5,75%	0,5791 -3,48%	0,7596 +26,60%

Analisando os resultados apresentados na Tabela 13, observamos que abordagem BoVW-SPM com uso da divergência KL/ε obteve ganhos de 26,60% em relação a Cosseno com a abordagem BoVW$_J$, desta forma apresentando resultados promissores quando comparado com o trabalho de (JÉGOU; CHUM, 2012). Destacando que a medida Cosseno e Euclidiana utilizando a abordagem BoVW-SPM não obteve resultados superiores que a Cosseno com a abordagem BoVW$_J$, mostrando mais uma vez a superioridade da divergência. Vale mencionar que não foi testado diferentes parâmetros para abordagem BoVW-SPM, como por exemplo o tamanho do *codebook* ou a quantidade de níveis das pirâmides espaciais, pois a importância deste trabalho é o uso das funções de similaridade e não o processo de representação da imagem por um histograma.

7.4 Conclusão dos experimentos

Em geral, todos os experimentos que utilizavam a GID e/ou a KL como medida de similaridade retornaram bons resultados quando comparadas com as distâncias tradicionais, neste caso a Cosseno e a Euclidiana. Acredita-se que os resultados obtidos com a GID e com a KL estão de acordo com a observação de (BANERJEE et al., 2005) "*the divergence should capture the similarity properties desirable in the application, and need not necessarily depend on how the data was actually generated*". Também, as similaridades baseadas nas divergências (GID e KL) correspondem à hipótese de distribuição de *Poisson*, a qual é mais apropriada para a utilização do descritor de imagem.

Neste trabalho, acredita-se que os resultados promissores que a GID e a KL retornaram nos experimentos comparadas com as medidas tradicionais (Euclidiana e Cosseno) foram principalmente o fato de os dados serem estatísticos.

Capítulo 8

Conclusão

Os sistemas CBIR têm sido amplamente aplicados na análise e classificação de imagens. Entretanto, conforme apresentado no capítulo introdutório, os algoritmos atuais dos sistemas CBIR apresentam limitações. Visando minimizar estas limitações, diversos trabalhos têm utilizado estes sistemas com diferentes medidas de similaridade.

Nesta pesquisa foram utilizadas as divergências de Bregman KL e GID para computar a similaridade entre vetores de características que representam as imagens, apresentando tratamentos adequados que garantem o domínio de aplicação destas divergências.

Para avaliar o desempenho do sistema CBIR, inicialmente, foi apresentada uma comparação dos resultados de busca de imagens utilizando as funções GID, KL, Cosseno e Euclidiana, considerando diferentes abordagens para a caracterização. A partir das medidas de avaliação: precisão x revocação, MAP, nDCG e MP, observa-se que, ao utilizar as divergências de Bregman, o sistema propicia ganhos relevantes na maioria dos experimentos, quando comparada com as distâncias Cosseno e Euclidiana.

Concluindo o primeiro grupo de experimentos, o menor ganho obtido comparando o desempenho da GID em relação à distância Cosseno foi de 4% considerando todas as abordagens para a caracterização. O maior ganho foi de 86% na avaliação MAP com abordagem BSM. De forma geral, os resultados mostram que a GID superou a distância Cosseno em todos os métodos de avaliação com as diferentes abordagens de caracterização. A utilização da divergência KL proporciona um ganho de até 79%, na avaliação MAP, em relação aos resultados obtidos com as outras funções de similaridade analisadas, evidenciando sua superioridade em relação às distâncias tradicionais. Estes resultados confirmam que o uso das divergências KL e GID na etapa de similaridade torna o sistema CBIR mais eficaz que o uso da distância Euclidiana e Cosseno.

Considerando o domínio das divergências de Bregman (KL e GID) com caracterizações que podem conter dados nulos em suas coordenadas, apresentamos tratamentos adequados que possibilitam a utilização destas divergências, que são: KL/ε (ou GID/ε), KL/N e GID/D.

De forma geral a GID/D e GID/ε apresentaram bons resultados quando comparados

com a distância Cosseno, com a caracterização BoVW, mostrando que estes tratamentos podem ser uma boa escolha para o uso da função GID. Observou-se ainda, o impacto da escolha dos parâmetros ε e α. Definindo um valor adequado para estes parâmetros pode-se chegar a um ganho de 33% (utilizando GID/ε) e 51% (usando a GID/D) em relação a distância Cosseno.

Os tratamentos propostos (KL/ε, KL/N e GID/D) quando comparados com a distância Cosseno, com a caracterização BoVW-SPM, obtiveram um ganho mínimo de 23% e um ganho máximo de 86% utilizando a função KL/ε na avaliação MAP. Em média, todos os tratamentos apresentaram bons resultados, excedendo os resultados apresentados pela distância Euclidiana e Cosseno, mostrando que o uso das divergências com os tratamentos pode se tornar vantajoso, desde que os parâmetros estejam configurados adequadamente.

Da análise ainda concluímos que utilização da DB com o tratamento proposto pela TI não proporciona bons resultados e inviabiliza a utilização das divergências KL e GID. Assim, observa-se que os resultados obtidos utilizando as DB's com os tratamentos, relatado neste estudo, são melhores quando comparados aos obtidos utilizando a DB's com tratamento proposto na TI e as distâncias convencionais (Cosseno e a Euclidiana). O que demonstra a viabilidade e eficácia dos tratamentos apresentados.

Analisando os resultados alcançados podemos concluir que o uso das divergências de Bregman (KL e GID) com os tratamentos apresentados, na etapa de similaridade, tornam os sistemas CBIR mais eficientes do que o uso das distâncias tradicionais (Euclidiana e Cosseno).

Em trabalhos futuros, planeja-se conseguir enriquecer o estudo das funções de Bregman, considerando modelos de caracterização mais sofisticados e complexos com banco de dados maiores, como também aprofundar os estudos em similaridade via representante de classes.

Resultados parciais deste trabalho foram aceitos para publicação e apresentados no 21st IEEE *International Conference on Electronics Circuits and Systems* (IEEE-ICECS-2014) em Dezembro de 2014 na cidade de Marseille na França, de qualificação B1. Aprovamos também um resumo expandido no XIV Semana da Matemática e IV Semana da Estatística que aconteceu na cidade de Uberlândia-MG na Universidade Federal de Uberlândia em 2014. Além disso, parte dessa pesquisa foi apresentada no VIII Workshop de Teses e Dissertações em Ciência da Computação na Universidade Federal de Uberlândia, Uberlândia-MG, 2014.

Referências

ABOOD, Z. I.; MUHSIN, I. J.; TAWFIQ, N. J. Content-based image retrieval (cbir) using hybrid technique. **International Journal of Computer Applications**, v. 83, n. 12, p. 17–24, December 2013.

AMATO, G.; FALCHI, F.; GENNARO, C. On reducing the number of visual words in the bag-of-features representation. In: **VISAPP 2013 - Proceedings of the International Conference on Computer Vision Theory and Applications, Volume 1, Barcelona, Spain, 21-24 February, 2013.** [S.l.: s.n.], 2013. p. 657–662.

ANDALÓ, F. A. et al. Shape feature extraction and description based on tensor scale. **Pattern Recognition**, v. 43, n. 1, p. 26–36, 2010.

BALAN, A. G. R. et al. Integrando textura e forma para a recuperação de imagens por conteúdo. In: **In: IX Congresso Brasileiro de Informática em Saúde**. Ribeirão Preto, SP, Brasil: ACM, 2004.

BAMPIS, L. et al. Real-time indexing for large image databases: color and edge directivity descriptor on GPU. **The Journal of Supercomputing**, v. 71, n. 3, p. 909–937, 2015.

BANERJEE, A. et al. Clustering with bregman divergences. **Journal of Machine Learning Research**, v. 6, p. 1705–1749, 2005.

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.

BUGATTI, P. H. **Arcabouço para recuperação de imagens por conteúdo visando à percepção do usuário.** Tese (Doutorado) — Universidade de São Paulo, 2012.

BUGATTI, P. H.; TRAINA, A. J. M.; JR., C. T. Assessing the best integration between distance-function and image-feature to answer similarity queries. In: WAINWRIGHT, R. L.; HADDAD, H. (Ed.). **SAC**. ACM, 2008. p. 1225–1230. Disponível em: <http://dblp.uni-trier.de/db/conf/sac/sac2008.html#BugattiTT08>.

CANNY, J. A computational approach to edge detection. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 8, n. 6, p. 679–698, jun. 1986. Disponível em: <http://dx.doi.org/10.1109/TPAMI.1986.4767851>.

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>.

CENSOR, Y. A.; ZENIOS, S. A. **Parallel Optimization: Theory, Algorithms and Applications**. [S.l.]: Oxford University Press, 1997. ISBN 019510062X.

CHANG, S. K.; HSU, A. Image information systems: Where do we go from here? **IEEE Trans. on Knowl. and Data Eng.**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 4, n. 5, p. 431–442, out. 1992.

COVER, T. M.; THOMAS, J. A. **Elements of Information Theory**. New York, NY, USA: Wiley-Interscience, 1991. ISBN 0-471-06259-6.

CSISZÁR. **Information geometry and alternating minimization procedures. Statistics and Decisions**. [S.l.: s.n.], 1995.

CSISZÁR, I. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. **The Annals of Statistics**, v. 19, n. 4, p. 2032–2066, 1991.

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.

DAS, M.; MANMATHA, R.; RISEMAN, E. M. Indexing flower patent images using domain knowledge. **IEEE Intelligent Systems**, v. 14, n. 5, p. 24–33, 1999. Disponível em: <http://dblp.uni-trier.de/db/journals/expert/expert14.html#DasMR99>.

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>.

DESERNO, T. M.; WELTER, P.; HORSCH, A. Towards a repository for standardized medical image and signal case data annotated with ground truth. **J. Digital Imaging**, v. 25, n. 2, p. 213–226, 2012. Disponível em: <http://dblp.uni-trier.de/db/journals/jdi/jdi25.html#DesernoWH12>.

DO, M. N.; VETTERLI, M. Texture similarity measurement using kullback-leibler distance on wavelet subbands. In: . [S.l.: s.n.], 2000. p. 730–733.

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>.

DUPRET, G.; PIWOWARSKI, B. Model based comparison of discounted cumulative gain and average precision. **J. Discrete Algorithms**, v. 18, p. 49–62, 2013. Disponível em: <http://dx.doi.org/10.1016/j.jda.2012.10.002>.

ELKAN, C. Using the Triangle Inequality to Accelerate k-Means. In: FAWCETT, T.; MISHRA, N. (Ed.). **Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA**. [S.l.]: AAAI Press, 2003. p. 147–153.

FENG, D.; SIU, W.; ZHANG, H. **Multimedia Information Retrieval and Management: Technological Fundamentals and Applications**. Springer, 2003. 3 p. (Engineering online library). Disponível em: <http://books.google.com.br/books?id=qn54Qv35sm8C>.

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>.

FLORINDO, J. B.; BRUNO, O. M. Fractal descriptors based on fourier spectrum applied to texture analysis. **CoRR**, abs/1201.4597, 2012. Disponível em: <http://arxiv.org/abs/1201.4597>.

FORGY, E. Cluster analysis of multivariate data: Efficiency versus interpretability of classification. **Biometrics**, v. 21, n. 3, p. 768–769, 1965.

FRINTROP, S.; ROME, E.; CHRISTENSEN, H. I. Computational visual attention systems and their cognitive foundations: A survey. **ACM Trans. Appl. Percept.**, ACM, New York, NY, USA, v. 7, n. 1, p. 6:1–6:39, jan. 2010. Disponível em: <http://doi.acm.org/10.1145/1658349.1658355>.

GODIL, A.; LIAN, Z.; WAGAN, A. Exploring local features and the bag-of-visual-words approach for bioimage classification. In: **ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics. ACM-BCB 2013, Washington, DC, USA, September 22-25, 2013**. [s.n.], 2013. p. 694. Disponível em: <http://doi.acm.org/10.1145/2506583.2512370>.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.

GRAUMAN, K.; DARRELL, T. The pyramid match kernel: Discriminative classification with sets of image features. In: **In ICCV**. [S.l.: s.n.], 2005. p. 1458–1465.

GREENSPAN, H. et al. Overcomplete steerable pyramid filters and rotation invariance. In: **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 1994. p. 222–228.

GRUNDLAND, M.; DODGSON, N. A. Decolorize: Fast, contrast enhancing, color to grayscale conversion. **Pattern Recognition**, v. 40, n. 11, p. 2891–2896, 2007. Disponível em: <http://dblp.uni-trier.de/db/journals/pr/pr40.html#GrundlandD07>.

GRUNWALD, P. D.; DAWID, A. P. **Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory**. 2004.

GRZESZICK, R.; ROTHACKER, L.; FINK, G. A. Bag-of-features representations using spatial visual vocabularies for object classification. In: **IEEE International Conference on Image Processing, ICIP 2013, Melbourne, Australia, September 15-18, 2013**. [s.n.], 2013. p. 2867–2871. Disponível em: <http://dx.doi.org/10.1109/ICIP.2013.6738590>.

HAREL, J.; KOCH, C.; PERONA, P. Graph-based visual saliency. In: **Advances in Neural Information Processing Systems 19**. [S.l.]: MIT Press, 2007. p. 545–552.

IMRAN, M.; HASHIM, R.; KHALID, N. E. A. Segmentation-based fractal texture analysis and color layout descriptor for content based image retrieval. In: **14th International Conference on Intelligent Systems Design and Applications, ISDA 2014, Okinawa, Japan, November 28-30, 2014**. [S.l.: s.n.], 2014. p. 30–33.

IQBAL, T. et al. Performance evaluation of content based image retrieval using indexed views. **International Journal of Computer, Information, Systems and Control Engineering**, World Academy of Science, Engineering and Technology, v. 8, n. 7, p. 1065 – 1068, 2014.

ITTI, L.; KOCH, C. Computational modelling of visual attention. **Nature Reviews Neuroscience**, v. 2, n. 3, p. 194–203, Mar 2001.

ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 20, n. 11, p. 1254–1259, Nov 1998.

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>.

JANG, K.; HAN, S.; KIM, I. Person re-identification based on color histogram and spatial configuration of dominant color regions. **CoRR**, abs/1411.3410, 2014. Disponível em: <http://arxiv.org/abs/1411.3410>.

JÄRVELIN, K.; KEKÄLÄINEN, J. Cumulated gain-based evaluation of ir techniques. **ACM Trans. Inf. Syst.**, ACM, New York, NY, USA, v. 20, n. 4, p. 422–446, out. 2002. Disponível em: <http://doi.acm.org/10.1145/582415.582418>.

JÉGOU, H.; CHUM, O. Negative evidences and co-occurrences in image retrieval: the benefit of PCA and whitening. In: **ECCV - European Conference on Computer Vision**. Firenze, Italy: [s.n.], 2012. Disponível em: <https://hal.inria.fr/hal-00722622>.

J.HERRMANN, F.; P.FRIEDLANDER, M.; YILMAZ, O. Fighting the curse of dimensionality: compressive sensing in exploration seismology. **IEEE SIgnal Processing Magazine**, v. 29, p. 88–100, May 2012.

KAZAKOS, D. Spectral distance measures between continuous-time vector gaussian processes (corresp.). **IEEE Trans. Inf. Theor.**, IEEE Press, Piscataway, NJ, USA, v. 28, n. 4, p. 679–681, set. 2006. ISSN 0018-9448. Disponível em: <http://dx.doi.org/10.1109/TIT.1982.1056521>.

KEKRE, H. B.; SONAWANE, K. Effect of similarity measures for cbir using bins approach. **International Journal of Image Processing (IJIP)**, v. 6, p. 182–197, 2012.

KIMURA, P. A. S. et al. Evaluating retrieval effectiveness of descriptors for searching in large image databases. **JIDM**, v. 2, n. 3, p. 305–320, 2011. Disponível em: <http://dblp.uni-trier.de/db/journals/jidm/jidm2.html#KimuraCSTG11>.

KINGSBURY, N. The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters. In: **in Proceedings of the 8th IEEE DSP Workshop, Bryce Canyon, Utah, USA**. [S.l.: s.n.], 1998. p. 9–12.

_____. Complex wavelets for shift invariant analysis and filtering of signals. **Applied and Computational Harmonic Analysis**, v. 10, n. 3, p. 234–253, maio 2001.

KOCH, C.; ULLMAN, S. Shifts in selective visual attention: towards the underlying neural circuitry. **Human Neurobiology**, v. 4, p. 219–227, 1985.

KRISHNAMOORTHY, K. **Handbook of Statistical Distributions with Applications**. [S.l.: s.n.], 2006.

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>.

KULIS, B.; SUSTIK, M.; DHILLON, I. Learning low-rank kernel matrices. In: **In ICML**. [S.l.]: Morgan Kaufmann, 2006. p. 505–512.

KWITT, R.; UHL, A. Image similarity measurement by kullback-leibler divergences between complex wavelet subband statistics for texture retrieval. In: **Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA**. [S.l.: s.n.], 2008. p. 933–936.

LADES, M. et al. Distortion invariant object recognition in the dynamic link architecture. **IEEE Trans. Comput.**, IEEE Computer Society, Washington, DC, USA, v. 42, n. 3, p. 300–311, mar. 1993.

LASMAR, N.; BERTHOUMIEU, Y. Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. **IEEE Transactions on Image Processing**, v. 23, n. 5, p. 2246–2261, 2014. Disponível em: <http://dx.doi.org/10.1109/TIP.2014.2313232>.

LAZEBNIK, S.; SCHMID, C.; PONCE, J. A sparse texture representation using local affine regions. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 27, p. 1265–1278, 2005.

_____. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: **Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2**. Washington, DC, USA: IEEE Computer Society, 2006. (CVPR '06), p. 2169–2178. Disponível em: <http://dx.doi.org/10.1109/CVPR.2006.68>.

LEGUAY, J.; FRIEDMAN, T.; CONAN, V. Dtn routing in a mobility pattern space. In: **Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking**. New York, NY, USA: ACM, 2005. (WDTN '05), p. 276–283.

LI, B. et al. Rank-sift: Learning to rank repeatable local interest points. In: **The 24th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, CO, USA, 20-25 June 2011**. [s.n.], 2011. p. 1737–1744. Disponível em: <http://dx.doi.org/10.1109/CVPR.2011.5995461>.

LIBERTY, E.; SRIHARSHA, R.; SVIRIDENKO, M. An algorithm for online k-means clustering. **CoRR**, abs/1412.5721, 2014. Disponível em: <http://arxiv.org/abs/1412.5721>.

LIU, H. et al. Comparing dissimilarity measures for content-based image retrieval. In: **Proceedings of the 4th Asia Information Retrieval Conference on Information Retrieval Technology**. Berlin, Heidelberg: Springer-Verlag, 2008. (AIRS'08), p. 44–50.

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>.

LONCARIC, S. A survey of shape analysis techniques. **Pattern Recognition**, v. 31, n. 8, p. 983–1001, 1998. Disponível em: <http://dblp.uni-trier.de/db/journals/pr/pr31.html#Loncaric98>.

LOWE, D. G. Object recognition from local scale-invariant features. In: **Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2**. Washington, DC, USA: IEEE Computer Society, 1999. (ICCV '99), p. 1150–. Disponível em: <http://dl.acm.org/citation.cfm?id=850924.851523>.

_____. 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>.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. New York, NY, USA: Cambridge University Press, 2008. ISBN 0521865719, 9780521865715.

MARQUES, P. M. A. et al. Recuperação de imagem baseada em conteúdo: Uso de atributos de textura para caracterização de microcalcificações mamográficas. **Radiologia Brasileira**, v. 35, p. 93–98, 2002.

MATHIASSEN, J. R.; SKAVHAUG, A.; Bø, K. Texture similarity measure using kullback-leibler divergence between gamma distributions. In: HEYDEN, A. et al. (Ed.). **ECCV (3)**. [S.l.]: Springer, 2002. (Lecture Notes in Computer Science, v. 2352), p. 133–147. ISBN 3-540-43746-0.

MCLACHLAN, G. Mahalanobis distance. **Resonance**, Springer India, v. 4, n. 6, p. 20–26, 1999.

MENDOZA-MARTINEZ, C.; ORTEGA, J. C. P.; ARREGUIN, J. M. R. A novel approach for face authentication using speeded up robust features algorithm. In: **Human-Inspired Computing and Its Applications - 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16-22, 2014. Proceedings, Part I**. [s.n.], 2014. p. 356–367. Disponível em: <http://dx.doi.org/10.1007/978-3-319-13647-9_33>.

MIKOLAJCZYK, K.; SCHMID, C. A performance evaluation of local descriptors. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 27, n. 10, p. 1615–1630, out. 2005. Disponível em: <http://dx.doi.org/10.1109/TPAMI.2005.188>.

MÜLLER, H. et al. **A Review of Content-Based Image Retrieval Systems in Medical Applications – Clinical Benefits and Future Directions**. 2004.

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.

NGUYEN, H. Q. et al. 3d human face recognition using sift descriptors of face's feature regions. In: **New Trends in Computational Collective Intelligence**. [s.n.], 2015. p. 117–126. Disponível em: <http://dx.doi.org/10.1007/978-3-319-10774-5_11>.

NIEBUR, E.; KOCH, C. Control of selective visual attention: Modeling the "where" pathway. In: TOURETZKY, D. S.; MOZER, M. C.; HASSELMO, M. E. (Ed.). **Advances in Neural Information Processing Systems**. Cambridge, MA: MIT Press, 1996. v. 8, p. 802–808.

_____. Computational architectures for attention. In: PARASURAMAN, R. (Ed.). **The Attentive Brain**. Cambridge, MA: MIT Press, 1998. cap. 9, p. 163–186.

OJALA, T.; PIETIKÄINEN, M.; HARWOOD, D. A comparative study of texture measures with classification based on featured distributions. **Pattern Recognition**, v. 29, n. 1, p. 51–59, jan. 1996.

PALUS, M. On entropy rates of dynamical systems and gaussian processes. **Physics Letters A**, v. 227, p. 301–308, 1997.

PEDROSA, G. V.; TRAINA, A. J. M.; JR., C. T. Using sub-dictionaries for image representation based on the bag-of-visual-words approach. In: **2014 IEEE 27th International Symposium on Computer-Based Medical Systems, New York, NY, USA, May 27-29, 2014**. [s.n.], 2014. p. 165–168. Disponível em: <http://dx.doi.org/10.1109/CBMS.2014.60>.

PENG, X. et al. Fast low-rank representation based spatial pyramid matching for image classification. **CoRR**, abs/1409.5786, 2014. Disponível em: <http://arxiv.org/abs/1409.5786>.

PIRO, P. et al. Image retrieval via kullback-leibler divergence of patches of multiscale coefficients in the KNN framework. In: **International Workshop on Content-Based Multimedia Indexing, CBMI 2008, London, UK, June 18-20, 2008**. [s.n.], 2008. p. 230–235. Disponível em: <http://dx.doi.org/10.1109/CBMI.2008.4564951>.

RAFIEE, G.; DLAY, S. S.; WOO, W. L. Region-of-interest extraction in low depth of field images using ensemble clustering and difference of gaussian approaches. **Pattern Recognition**, v. 46, n. 10, p. 2685–2699, 2013. Disponível em: <http://dx.doi.org/10.1016/j.patcog.2013.03.006>.

RAJALAKSHMI, M.; SUBASHINI, P. Texture based image segmentation of chili pepper x-ray images using gabor filter. **CoRR**, abs/1405.1966, 2014. Disponível em: <http://arxiv.org/abs/1405.1966>.

RAJASHEKAR, U. et al. Gaffe: A gaze-attentive fixation finding engine. **IEEE Transactions on Image Processing (TIP)**, v. 17, n. 4, p. 564–573, 2008. Disponível em: <http://dx.doi.org/10.1109/TIP.2008.917218>.

RENALS, S. et al. Indexing and retrieval of broadcast news. **Speech Communication**, v. 32, n. 1-2, p. 5–20, 2000. Disponível em: <http://dblp.uni-trier.de/db/journals/speech/speech32.html#RenalsAKR00>.

ROCHA, B. M. et al. Image retrieval via generalized i-divergence in the bag-of-visual-words framework. In: **21st IEEE International Conference on Electronics, Circuits and Systems, ICECS 2014, Marseille, France, December 7-10, 2014**. [S.l.: s.n.], 2014. p. 734–737.

ROCKAFELLAR, R. T. **Convex Analysis (Princeton Landmarks in Mathematics and Physics)**. [S.l.]: Princeton University Press, 1996. Paperback.

ROHANI, B.; NUGOHO, B. Manhattan-chebychev distance metric for mimo systems. **Technical report of IEICE. RCS**, The Institute of Electronics, Information and Communication Engineers, v. 108, n. 305, p. 49–52, nov 2008. ISSN 09135685. Disponível em: <http://ci.nii.ac.jp/naid/110007100559/en/>.

SAKJI-NSIBI, S.; BENAZZA-BENYAHIA, A. Fast scalable retrieval of multispectral images with kullback-leibler divergence. In: **ICIP**. IEEE, 2010. p. 2333–2336. ISBN 978-1-4244-7994-8. Disponível em: <http://dblp.uni-trier.de/db/conf/icip/icip2010.html#Sakji-NsibiB10>.

SANTINI, S.; JAIN, R. Similarity measures. **IEEE Trans. Pattern Anal. Mach. Intell.**, IEEE Computer Society, Washington, DC, USA, v. 21, n. 9, p. 871–883, set. 1999. ISSN 0162-8828. Disponível em: <http://dx.doi.org/10.1109/34.790428>.

SCHOLAR, K. M. T. Image similarity measure using color histogram, color coherence vector, and sobel method. In: **Internatiional Journal of Science and Research (IJSR)**. [S.l.: s.n.], 2013. p. 538–543.

SCHWANDER, O.; NIELSEN, F. Reranking with contextual dissimilarity measures from representational bregman k-means. In: RICHARD, P.; BRAZ, J. (Ed.). **VISAPP (1)**. INSTICC Press, 2010. p. 118–123. Disponível em: <http://dblp.uni-trier.de/db/conf/visapp/visapp2010-1.html#SchwanderN10>.

SHILANE, P. et al. The Princeton shape benchmark. In: **Shape Modeling International**. [S.l.: s.n.], 2004.

SIDRAM, M. H.; BHAJANTRI, N. U. Enhancement of mean shift tracking through joint histogram of color and color coherence vector. In: **Proceedings of the Second International Conference on Soft Computing for Problem Solving, SocProS 2012, December 28-30, 2012, JK Lakshmipat University (JKLU), Jaipur, India**. [S.l.: s.n.], 2012. p. 547–555.

SILVA, M. P. **Texture Based Image Segmentation of Chili Pepper X-Ray Images Using Gabor Filter**. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (USP), 2014.

SIVIC, J.; ZISSERMAN, A. Video Google: A text retrieval approach to object matching in videos. In: **Proceedings of the International Conference on Computer Vision**. [s.n.], 2003. v. 2, p. 1470–1477. Disponível em: <http://www.robots.ox.ac.uk/~vgg>.

SNOEK, C. G. M.; SMEULDERS, A. W. M. Visual-concept search solved? **IEEE Computer**, v. 43, n. 6, p. 76–78, 2010. Disponível em: <http://www.science.uva.nl/research/publications/2010/SnoekIC2010>.

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>.

SPERTUS, E.; SAHAMI, M.; BUYUKKOKTEN, O. Evaluating similarity measures: A large-scale study in the orkut social network. In: **Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining**. New York, NY, USA: ACM, 2005. (KDD '05), p. 678–684.

SWAIN, M. J.; BALLARD, D. H. Color indexing. **International Journal of Computer Vision**, v. 7, p. 11–32, 1991.

TAMURA, H.; YOKOYA, N. Image database systems: A survey. **Pattern Recognition**, v. 17, n. 1, p. 29–43, 1984.

TORRES, R. da S.; FALCÃO, A. X. Content-based image retrieval: Theory and applications. **RITA**, v. 13, n. 2, p. 161–185, 2006. Disponível em: <http://dblp.uni-trier.de/db/journals/rita/rita13.html#TorresF06>.

_____. Contour salience descriptors for effective image retrieval and analysis. **IMAGE AND VISION COMPUTING**, v. 25, p. 3–13, 2007.

TSOTSOS, J. Analyzing vision at the complexity level. **Behavioral and Brain Sciences**, v. 13, n. 3, p. 423–469, 1990.

ULLMAN, S.; VIDAL-NAQUET, M.; SALI, E. Visual features of intermediate complexity and their use in classification. **Nature neuroscience**, v. 5, n. 7, p. 682–687, jul. 2002. ISSN 1097-6256. Disponível em: <http://dx.doi.org/10.1038/nn870>.

VALLE, E.; CORD, M. Advanced techniques in cbir: Local descriptors, visual dictionaries and bags of features. In: **Proceedings of the 2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing**. Washington, DC, USA: IEEE Computer Society, 2009. p. 72–78. ISBN 978-0-7695-3815-0. Disponível em: <http://dx.doi.org/10.1109/SIBGRAPI-Tutorials.2009.14>.

VASCONCELOS, N.; LIPPMAN, A. A unifying view of image similarity. In: **IEEE International Conference on Pattern Recognition**. [S.l.: s.n.], 2000. p. 1038–1041.

VEDALDI, A.; FULKERSON, B. Vlfeat: An open and portable library of computer vision algorithms. In: **Proceedings of the International Conference on Multimedia**. New York, NY, USA: ACM, 2010. (MM '10), p. 1469–1472. Disponível em: <http://doi.acm.org/10.1145/1873951.1874249>.

WANG, J. Joint-vivo: Selecting and weighting visual words jointly for bag-of-features based tissue classification in medical images. **CoRR**, abs/1208.3822, 2012. Disponível em: <http://arxiv.org/abs/1208.3822>.

WANG, J. et al. Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. **IEEE Trans. Med. Imaging**, v. 30, n. 11, p. 1996–2011, 2011. Disponível em: <http://dx.doi.org/10.1109/TMI.2011.2161673>.

WEN, Z.; ZHANG, R.; RAMAMOHANARAO, K. Enabling precision/recall preferences for semi-supervised SVM training. In: **Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014**. [s.n.], 2014. p. 421–430. Disponível em: <http://doi.acm.org/10.1145/2661829.2661977>.

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.

A Robust Similarity Measure Method in CBIR System, v. 2. 662–666 p.

ZHANG, D.; LU, G. Review of shape representation and description techniques. **Pattern Recognition**, v. 37, p. 1–19, 1 2004. Disponível em: <http://www.sciencedirect.com/science/article/B6V14-49PYM0K-1/2/2cda972cc0c485d0af580a8226ccd864>.

ZHAO, H.; WANG, H.; KHAN, M. K. Steganalysis for palette-based images using generalized difference image and color correlogram. **Signal Processing**, v. 91, n. 11, p. 2595–2605, 2011.

ZHOU, D. et al. Ranking on data manifolds. In: **Advances in Neural Information Processing Systems 16**. [S.l.]: MIT Press, 2004.

ZHOU, Z.-H.; DAI, H.-B. Query-sensitive similarity measure for content-based image retrieval. In: **ICDM**. [S.l.]: IEEE Computer Society, 2006. p. 1211–1215.

ZHU, L.; RAO, A. B.; ZHANG, A. Theory of keyblock-based image retrieval. **ACM Trans. Inf. Syst.**, ACM, New York, NY, USA, v. 20, n. 2, p. 224–257, abr. 2002. Disponível em: <http://doi.acm.org/10.1145/506309.506313>.

ZICKLER, S.; EFROS, A. A. Detection of multiple deformable objects using PCA-SIFT. In: **Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, July 22-26, 2007, Vancouver, British Columbia, Canada**. [s.n.], 2007. p. 1127–1133. Disponível em: <http://www.aaai.org/Library/AAAI/2007/aaai07-179.php>.

ZUIDERVELD, K. Graphics gems iv. In: HECKBERT, P. S. (Ed.). San Diego, CA, USA: Academic Press Professional, Inc., 1994. cap. Contrast Limited Adaptive Histogram Equalization, p. 474–485.