

---

Exploração visual do espaço de  
características: uma abordagem para  
análise de imagens via projeção de  
dados multidimensionais

*Bruno Brandoli Machado*

---



SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-  
USP

Data de Depósito:

Assinatura: \_\_\_\_\_

*Bruno Brandoli Machado*

**Orientador:** *Prof. Dr. João do Espírito Santo Batista Neto*

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP como parte dos requisitos necessários à obtenção do título de Mestre em Ciências de Computação e Matemática Computacional.

**USP - São Carlos**  
**Novembro/2010**



***Aos meus pais.***



# Agradecimentos

---

Agradeço primeiramente a Deus pela minha saúde e força durante esses anos. À minha família, especialmente meus pais Carlos e Mara Cristina, meus irmãos C.Nino e M.Popi, e minha avó Marleny pelo incentivo e amor que tiveram por mim. Ao Prof. João Batista pela oportunidade e orientação no Instituto de Ciências Matemáticas e de Computação (ICMC/USP); aos professores, funcionários e a todos que, direta- ou indiretamente, colaboraram para a realização deste trabalho. Eu também gostaria de agradecer aos membros do laboratório VICG/LCAD - Prof.(a) Rosane Minghim, Prof. Fernando Paulovich e Danilo Eler - pelas contribuições e valiosas discussões. Pela alegria na companhia dos colegas W.Hitoshi, Jorge Poco, Frizzi San Roman e Christian Wong. Aos amigos de pós-graduação e de república Rodrigo, Leandro, Jonathan e Wesley. Finalmente, eu gostaria de agradecer ao CNPq pela bolsa concedida durante o mestrado.





# Resumo

---

Sistemas para análise de imagens partem da premissa de que o conjunto de dados sob investigação está corretamente representado por características. Entretanto, definir quais características representam apropriadamente um conjunto de dados é uma tarefa desafiadora e exaustiva. Grande parte das técnicas de descrição existentes na literatura, especialmente quando os dados têm alta dimensionalidade, são baseadas puramente em medidas estatísticas ou abordagens baseadas em inteligência artificial, e normalmente são “caixas-pretas” para os usuários. A abordagem proposta nesta dissertação busca abrir esta “caixa-preta” por meio de representações visuais criadas pela técnica *Multidimensional Classical Scaling*, permitindo que usuários capturem interativamente a essência sobre a representatividade das características computadas de diferentes descritores. A abordagem é avaliada sobre seis conjuntos de imagens que contém texturas, imagens médicas e cenas naturais. Os experimentos mostram que, conforme a combinação de um conjunto de características melhora a qualidade da representação visual, a acurácia de classificação também melhora. A qualidade das representações é medida pelo índice da silhueta, superando problemas relacionados com a subjetividade de conclusões baseadas puramente em análise visual. Além disso, a capacidade de exploração visual do conjunto sob análise permite que usuários investiguem um dos maiores desafios em classificação de dados: a presença de variação intra-classe. Os resultados sugerem fortemente que esta abordagem pode ser empregada com sucesso como um guia para auxiliar especialistas a explorar, refinar e definir as características que representam apropriadamente um conjunto de imagens.

**Keywords:** *Exploração Visual do Espaço de Características; Análise Visual do Espaço de Características; Visualização de Espaços de Características; Avaliação do Espaço de Características*



# Abstract

---

---

Image analysis systems rely on the fact that the dataset under investigation is correctly represented by features. However, defining a set of features that properly represents a dataset is still a challenging and, in most cases, an exhausting task. Most of the available techniques, especially when a large number of features is considered, are based on purely quantitative statistical measures or approaches based on artificial intelligence, and normally are “black-boxes” to the user. The approach proposed here seeks to open this “black-box” by means of visual representations via *Multidimensional Classical Scaling* projection technique, enabling users to get insight about the meaning and representativeness of the features computed from different feature extraction algorithms and sets of parameters. This approach is evaluated over six image datasets that contains textures, medical images and outdoor scenes. The results show that, as the combination of sets of features and changes in parameters improves the quality of the visual representation, the accuracy of the classification for the computed features also improves. In order to reduce this subjectiveness, a measure called silhouette index, which was originally proposed to evaluate results of clustering algorithms, is employed. Moreover, the visual exploration of datasets under analysis enable users to investigate one of the greatest challenges in data classification: the presence of intra-class variation. The results strongly suggest that our approach can be successfully employed as a guidance to defining and understanding a set of features that properly represents an image dataset.

**Keywords:** *Visual Feature Space Exploration; Visual Feature Space Analysis; Feature Space Visualization; Feature Space Evaluation; Interactive Feature Space Exploration*



# Sumário

---

Sumário . . . . .	x
Lista de Figuras . . . . .	xi
Lista de Tabelas . . . . .	xiii
<b>1 INTRODUÇÃO</b>	<b>1</b>
1.1 Contextualização e Motivação . . . . .	1
1.2 Objetivos . . . . .	2
1.3 Organização do Texto . . . . .	3
<b>2 PROJEÇÃO DE DADOS MULTIDIMENSIONAIS</b>	<b>5</b>
2.1 Considerações Iniciais . . . . .	5
2.2 Técnicas de Projeção . . . . .	6
2.2.1 Sammon's Mapping . . . . .	7
2.2.2 Self-Organizing Map . . . . .	7
2.2.3 Interactive Document Map . . . . .	8
2.2.4 Least Square Projection . . . . .	8
2.2.5 Neighbor-Joining . . . . .	9
2.2.6 Multidimensional Scaling . . . . .	9
2.3 Avaliação da Qualidade da Projeção . . . . .	13
2.3.1 Índice da Silhueta . . . . .	14
2.4 Ferramenta PEx-Image . . . . .	15
2.5 Considerações Finais . . . . .	15
<b>3 DESCRIÇÃO DE IMAGENS</b>	<b>19</b>
3.1 Considerações Iniciais . . . . .	19
3.2 Técnicas para Descrição de Imagens . . . . .	20
3.3 Análise de Textura . . . . .	20
3.3.1 Matriz de Co-ocorrência . . . . .	21
3.3.2 Filtros de Gabor . . . . .	22
3.4 Análise de Cor . . . . .	23

3.4.1	Modelo de Reflectância em Constância de Cor . . . . .	23
3.4.2	Histogramas . . . . .	25
3.4.3	Momentos . . . . .	27
3.4.4	Scale Invariant Feature Transform - SIFT . . . . .	29
3.5	Características Locais e o Modelo Bag-of-Visual-Features . . . . .	30
3.6	Considerações Finais . . . . .	32
<b>4</b>	<b>UMA NOVA ABORDAGEM PARA EXPLORAÇÃO VISUAL DO ESPAÇO DE CARACTERÍSTICAS APLICADA À ANÁLISE DE IMAGENS</b>	<b>35</b>
4.1	Metodologia Proposta . . . . .	35
4.2	Avaliação Experimental . . . . .	38
4.2.1	Conjuntos de Imagens . . . . .	38
4.2.2	Configuração dos Experimentos . . . . .	42
4.2.3	Resultados . . . . .	44
4.3	Considerações Finais . . . . .	57
<b>5</b>	<b>CONCLUSÃO</b>	<b>61</b>
5.1	Limitações . . . . .	62
5.2	Contribuições . . . . .	62
5.3	Trabalhos Futuros . . . . .	63
	<b>Referências</b>	<b>76</b>

# Lista de Figuras

---

---

2.1	Um exemplo de projeção 2D do conjunto de imagens <i>Corel</i> (Li and Wang (2003)). O conjunto possui 10 classes de cenas naturais com 100 imagens por classe. Em (a) os elementos visuais são representados por pontos e em (b) a própria imagem é utilizada como rótulo. . . . .	6
2.2	Um exemplo de projeção das imagens do conjunto <i>Corel</i> com a técnica <i>Neighbor-Joining</i> . O vetor de característica é composto por características extraídas de histogramas de cor. Em (a) os elementos visuais são representados por círculos e em (b) a própria imagem é utilizada como rótulo. . . . .	10
2.3	Um exemplo de exploração visual do espaço de características com a ferramenta <i>PEX-Image</i> . Neste caso específico, o usuário usa recursos de interação para explorar as imagens dos grupos <i>pratos</i> e <i>flores</i> do conjunto <i>Corel</i> . . . . .	16
3.1	Fluxograma para construção do modelo <i>bag-of-visual-features</i> - (BoVF). . . . .	33
4.1	Proposta de exploração visual do espaço de características. . . . .	37
4.2	Exemplo de três imagens de texturas de Brodatz. Cada imagem é dividida em nove sub-imagens sem sobreposição. . . . .	39
4.3	Quatro exemplos de texturas do conjunto <i>Textured Surfaces</i> . Este conjunto supera críticas sobre <i>Brodatz</i> devido à variação em condições de iluminação, mudanças no ponto de vista da câmera, além da não planaridade nas superfícies de textura, como a classe <i>casca de árvore</i> . . . . .	40

4.4	Exemplos da classe <i>tecido</i> do conjunto de materiais <i>KTH-TIPS</i> . As mudanças de escala e iluminação aumentam a variação intra-classe. As duas linhas de imagens mostram a mudança de iluminação para o mesmo material. As colunas representam as nove escalas do conjunto. . . . .	40
4.5	Exemplos de imagens do conjunto <i>ImageCLEF2006</i> . Este conjunto é caracterizado pela alta variação intra-classe. . . . .	41
4.6	Exemplos de imagens do conjunto <i>Corel</i> separadas em 10 classes. Apesar da presença de variação intra-classe, este conjunto possui alta correlação em termos de fundo. . . . .	41
4.7	Dois exemplos para cada classe de imagens do conjunto <i>Caltech</i> . Ao contrário do conjunto do <i>Corel</i> , este banco possui classes em que a informação de fundo não é correlata. . . . .	42
4.8	Análise de características de Gabor com 7 classes de texturas do álbum de <i>Brodatz</i> . Cinco classes de imagens são bem separadas, mas 2 são misturadas entre si. Isto indica que características extraídas não representam efetivamente este conjunto de imagens. . . . .	45
4.9	Projeção de 10 classes de texturas de <i>Brodatz</i> usando características de Gabor. Devido a baixa uniformidade no padrão da classe mármore, é facilmente verificado, mesmo por um usuário não-especialista ou sem conhecimento <i>a priori</i> , que o conjunto de características falha ao representar as imagens. Isso fornece ao especialista novas hipóteses de exploração do espaço. . . . .	46
4.10	Características de Gabor usando apenas a orientação $90^\circ$ . Duas das classes de imagens previamente adicionadas agora estão separadas, mas a coesão de todas as outras classes é reduzida; um resultado difícil de “ver” sem o apoio de visualização. . . . .	47
4.11	Projeção das classes selecionadas de <i>Broadtz</i> usando características extraídas pela técnica matriz de co-ocorrência. As duas classes de linho, não devidamente separadas pelas características de Gabor, indicam que a matriz de co-ocorrência é mais adequada para este conjunto de classes. . . . .	48
4.12	Projeção das texturas de <i>Brodatz</i> usando características computadas pela técnica matriz de co-ocorrência, sem a medida de energia. A projeção resultante apresenta tanto a melhor qualidade visual, quanto o maior valor do índice da silhueta. Isto indica que as características de energia são “ruidosas” para esse conjunto de dados. . . . .	49



4.13	Relação entre o índice da silhueta e taxa de classificação para três diferentes classificadores. As curvas do índice da silhueta e o desempenho de classificação apresentam a mesma tendência, isto é, quando a silhueta aumenta a classificação aumenta, e vice-versa. . . . .	51
4.14	Avaliação do conjunto <i>KTH-TIPS</i> influenciado pelas distâncias da técnica matriz de co-ocorrência. . . . .	53
4.15	Comparação de diferentes conjuntos de características computados de técnicas tradicionais de análise de textura sobre o conjunto <i>KTH-TIPS</i> . Novamente as taxas de classificação correspondem aos índices de silhueta, mostrando a generalidade desta abordagem. . . . .	54
4.16	Projeções de espaços de características diferentes extraídas do conjunto de imagens <i>KTH-TIPS</i> . Apesar de PCA ter sido aplicado para reduzir a dimensionalidade, as melhores taxas de classificação e os valores de silhueta condizem com a separabilidade de classe observada na projeção. . . . .	58
4.17	Resultados para tamanhos de vocabulários diferentes do modelo BoVF, considerando 50.000 e 100.000 <i>keypoints</i> amostrados sobre o conjunto de imagens <i>Corel</i> . . . . .	59
4.18	Resultados para diferentes tamanhos de vocabulário do modelo BoVF considerando 50.000 e 100.000 <i>keypoints</i> para o conjunto <i>Caltech</i> . . . . .	59
4.19	Comparação das projeções de dois espaços de características. A melhor projeção, em termos de separação de classes, coincide com as taxas de classificação, indicando que as projeções podem ser ferramentas úteis para apoiar os especialistas. . . . .	60



# Lista de Tabelas

---

---

2.1 Principais técnicas de projeção, seus respectivos autores e custos computacionais. . . . .	17
3.1 Medidas estatísticas calculadas por meio das matrizes de co-ocorrências. A primeira coluna da tabela contém o nome das medidas e a segunda coluna apresenta a formalização da referida medida. . . . .	21
4.1 As quatro primeiras colunas da tabela mostram as ponderações para as características de Gabor. Cada parametrização gera uma projeção diferente, avaliadas pelo índice da silhueta. A última coluna sustenta a hipótese que os resultados visuais condizem com o desempenho de classificação. . . . .	50
4.2 As cinco primeiras linhas correspondem às combinações das características de Gabor e co-ocorrência. O índice da silhueta calculado pelas projeções resultantes e as taxas de classificação são mostradas nas últimas duas colunas, respectivamente. . . . .	50
4.3 Análise dos índices de silhueta e taxas de classificação para três diferentes conjuntos de características. Os resultados obtidos considerando o conjunto de treinamento é preservado quando o conjunto de imagens inteiro é considerado. . . . .	52



# Lista de Abreviaturas

---

PEx	<i>Projection Explorer</i>
PEx-Image	<i>Projection Explorer for Images</i>
MDS	<i>Multidimensional Scaling</i>
CS	<i>Classical Scaling</i>
LMDS	<i>Landmark Multidimensional Scaling</i>
PCA	<i>Principal Components Analysis</i>
NJ	<i>Neighbor-Joining</i>
SOM	<i>Self-Organizing Map</i>
BMU	<i>Best Matching Unit</i>
SM	<i>Sammon's Mapping</i>
IDMAP	<i>Interactive Document Map</i>
ProjClus	<i>Projection by Clustering</i>
LSP	<i>Least Square Projection</i>
PPS	<i>Projection Precision Score</i>
LoG	<i>Laplacian of Gaussian</i>
DoG	<i>Difference of Gaussians</i>
RGB	<i>Red-Green-Blue</i>
HSI	<i>Hue-Saturation-Intensity</i>
SIFT	<i>Scale Invariant Feature Transform</i>
BoVF	<i>Bag of Visual Features</i>
NN	<i>Nearest Neighbors</i>
SVM	<i>Support Vector Machine</i>
ROC	<i>Receiver Operating Characteristic</i>



# Lista de Artigos

---

---

- BRANDOLI, B. M., ELER, D. M., PAULOVICH, F.V., MINGHIM, R., NETO, J.E.S.B. Visual Data Exploration to Feature Space Definition. In: *Proceedings of 23rd SIBGRAPI - Conference on Graphics, Patterns and Images*, Gramado, RS. Los Alamitos, CA, IEEE Computer Society, 2010. v. 1. p. 32-39.
- SILVA, S.; BRANDOLI, B. M., ELER, D. M., J.E.S.B, TRAINA, A.J. Silhouette-based Feature Selection for Classification of Medical Images. In: *The 23rd IEEE International Symposium on Computer-Based Medical Systems*, Perth, Australia. IEEE Computer Society, 2010. p. 315-320.
- SALAZAR, F.S.R; BRANDOLI, B. M., OCSA, A., OLIVEIRA, M.C.F. Cloth simulation using AABB hierarchical and Parallelism via GPU. In: *IX Simpósio Brasileiro de Jogos e Entretenimento Digital*, Florianópolis, SC. ISSN: 2179-2259. Sociedade Brasileira de Computação, 2010. p. 12-20.
- BRANDOLI, B. M., ELER, D. M., PAULOVICH, F.V., MINGHIM, R., NETO, J.E.S.B. Visual Data Analysis to Effective Feature Space Definition. In: *24th Annual ACM Symposium on Applied Computing*, Sierre, Switzerland. ACM, 2009.
- GONCALVES, W.N., BRANDOLI, B. M., NETO, J.E.S.B., BRUNO, O.M. A Complex Network Approach to Texture Applied to Medical Image Classification. In: *II ECCOMAS Thematic Conferences on Computational Vision and Medical Image Processing*, Porto, Portugal, 2009. p.32-40.





---

# INTRODUÇÃO

---

## 1.1 Contextualização e Motivação

Computar vetores de características é uma etapa crucial em diversos sistemas de análise de imagens, presentes em *visão computacional*, *recuperação baseada em conteúdo* e *reconhecimento de padrão*. Estes vetores são fundamentais para determinar a acurácia alcançada em aplicações dessa natureza (Theodoridis and Koutroumbas (2006)). Não somente existem diversas técnicas para computar tais vetores, como também são consideradas as tarefas de pré-processamento e normalização, além de um conjunto próprio de parâmetros cujo intuito é capturar diferentes propriedades das imagens. Por estes motivos, definir quais conjuntos de características fornecem alta taxa de classificação ou recuperação mais eficiente é considerada uma tarefa difícil (Pampalk et al. (2003); Schreck et al. (2008, 2010)). Em muitos casos, exige-se do pesquisador conhecimento *a priori* sobre o conjunto de dados. Além disso, diversos experimentos podem ser necessários para a construção de modelos robustos de classificação e, frequentemente, isto é um processo computacionalmente caro.

Geralmente, usuários-especialistas começam selecionando um conjunto pré-classificado de imagens e definem os parâmetros para as técnicas de extração (descritores), muitas vezes, uma tarefa trabalhosa e que demanda muito tempo. Após computar os vetores de características, tem início a classificação. Altas taxas de classificação serão somente alcançadas se o conjunto de características representa adequadamente o conjunto de imagens. Isto pode ser avaliado somente após o término do processo de classificação. Em outras pa-

lavras, é impossível inferir antecipadamente o quão efetivas as características serão no desempenho de classificação. Assim, diversas técnicas e ferramentas de visualização foram desenvolvidas para permitir que usuários explorem dados (Keim et al. (2008); Schreck et al. (2008)), algumas delas especialmente projetadas para explorar espaços multidimensionais resultantes da etapa de extração de características.

Uma destas abordagens foi proposta por Rodrigues Jr. et al. (2003) (depois estendida em Rodrigues Jr. et al. (2005)). Seu objetivo é apoiar a análise de características empregadas em buscas de similaridades para um sistema de recuperação de imagens. Uma vez que o vetor de características é computado, uma representação visual é criada, mostrando que a melhor representação obtém as melhores medidas de precisão e revocação quando uma consulta é executada sobre os dados. Entretanto, a representação visual não é usada para ajudar o usuário, a interativamente, definir o melhor conjunto de características ou o melhor conjunto de parâmetros, mas apenas confirma que o valor da medida de precisão e revocação condiz com a qualidade da representação visual. Similar ao trabalho de Rodrigues Jr. et al. (2003), de Leeuw and van Liere (2003) propuseram uma abordagem para apoiar projetistas de sistemas de recuperação de imagens. Embora representações visuais sejam criadas para que usuários possam interagir com o espaço de características, de Leeuw and van Liere (2003) não empregam medidas objetivas para ajudar usuários avaliarem a qualidade da projeção.

*PEX-Image* é uma ferramenta similar que cria representações visuais para explorar coleções de imagens (Eler et al. (2009)). Ela fornece visualizações interativas para auxiliar a exploração de espaços de características, e apoia a comparação entre diferentes espaços usando técnicas de coordenação. Esta ferramenta tem por objetivo criar a melhor representação visual para um conjunto de imagens, desconsiderando o fato que a representação deve refletir as estruturas existentes no espaço original. Dessa maneira, a abordagem proposta por Eler et al. (2009) fornece uma visão pouco precisa das relações de similaridade entre as imagens. Assim como de Leeuw and van Liere (2003), a qualidade das projeções criadas na ferramenta *PEX-Image* é somente avaliada de acordo com o ponto de vista do usuário.

## 1.2 Objetivos

O objetivo deste projeto de mestrado é criar uma abordagem para análise visual do espaço de características usando representações gráficas produzidas por técnicas de projeção. Em contraste com as abordagens apresentadas, o presente trabalho utiliza tais representações como um guia interativo para

explorar, definir e refinar conjuntos de características, dando entendimento sobre como os parâmetros das técnicas de descrição de imagens ou ponderações afetam as relações de similaridade entre grupos ou imagens individuais. A qualidade de tais representações é avaliada pelo *índice da silhueta* (Vendramin et al. (2009)), assim reduzindo a subjetividade de conclusões baseadas puramente em análise visual. Além disso, a abordagem proposta neste trabalho busca criar uma representação visual que melhor preserva a informação estrutural dos dados. Para este fim é usada a técnica de projeção *Multidimensional Classical Scaling* (Cox and Cox (2001)).

### 1.3 Organização do Texto

Esta dissertação está organizada da seguinte maneira:

- No Capítulo 2 é apresentada uma revisão das *técnicas de projeção* mais relevantes para o contexto deste trabalho. Nesse capítulo também são apresentadas medidas usadas para avaliar a qualidade das projeções e discutidos os critérios usados em *análise de cluster*. Por fim, o capítulo descreve a ferramenta de visualização *PEx-Image* sobre a qual este trabalho foi desenvolvido.
- No Capítulo 3 são apresentados dois grupos de técnicas para *descrição de imagens*, divididos mediante as propriedades físicas de *textura* e *cor*. O primeiro grupo descreve métodos tradicionais em *análise de textura* e o segundo descreve as principais técnicas de cor, incluindo o estado-da-arte em *análise de cor*. Por fim, são discutidas as estratégias *global* e *local* para representação de imagens.
- No Capítulo 4 é proposta uma abordagem para *exploração visual do espaço de características* aplicada à análise de imagens. Os conjuntos de imagens e os estudos de casos para avaliar esta proposta são também apresentados nesse capítulo.
- Finalmente, no Capítulo 5 são apresentadas as considerações finais para este trabalho, suas principais contribuições em exploração visual de espaços de características e propostas de trabalhos futuros.



---

# PROJEÇÃO DE DADOS MULTIDIMENSIONAIS

---

Este capítulo define o que é uma projeção de dados multidimensionais e em que contexto ela está inserida dentro da área de *visualização de informação*. A Seção 2.1 introduz o conceito básico de projeção de dados e, na Seção 2.2, é apresentada uma revisão das técnicas de visualização baseadas em projeção. Em seguida, na Seção 2.3, descreve-se como as projeções podem ser avaliadas quantitativamente. A ferramenta de visualização PEx-Image, sobre a qual este trabalho foi desenvolvido, é descrita na Seção 2.4. Por fim, as considerações finais do capítulo são feitas na Seção 2.5.

## 2.1 Considerações Iniciais

Uma importante classe de visualização exploratória de dados são as técnicas de projeção multidimensional. Elas são também estudadas em *estatística* e referidas como técnicas para projeção de dados multivariados ou gráficos estatísticos (Cleveland and McGill (1985); Grinstein et al. (1995); Wilkinson et al. (2006)). O objetivo de tais técnicas é criar representações visuais em baixa dimensionalidade para que usuários empreguem sua habilidade em reconhecer padrões e estruturas presentes no conjunto de dados. Basicamente, uma técnica de projeção mapeia dados  $m$ -dimensionais em um espaço  $p$ -dimensional, para  $p = \{1,2,3\}$ . Cada instância de dados é representada como um elemento visual na projeção, por meio de um círculo, ponto ou a própria imagem. A posição geométrica destes elementos reflete informações sobre a estrutura do conjunto de dados, tais como similaridades entre pares de instâncias, for-

mação de grupos e o relacionamento entre tais grupos. Assim, os elementos posicionados próximos uns aos outros na representação final indicam que as instâncias de dados que eles representam são similares com relação a uma certa medida de distância. Contrariamente, os elementos que estão distantes na projeção indicam que os objetos não são correlacionados de acordo com o conjunto de características explorado. As Figuras 2.1(a) e 2.1(b) mostram um exemplo de mapeamento usando pontos e imagens, respectivamente.

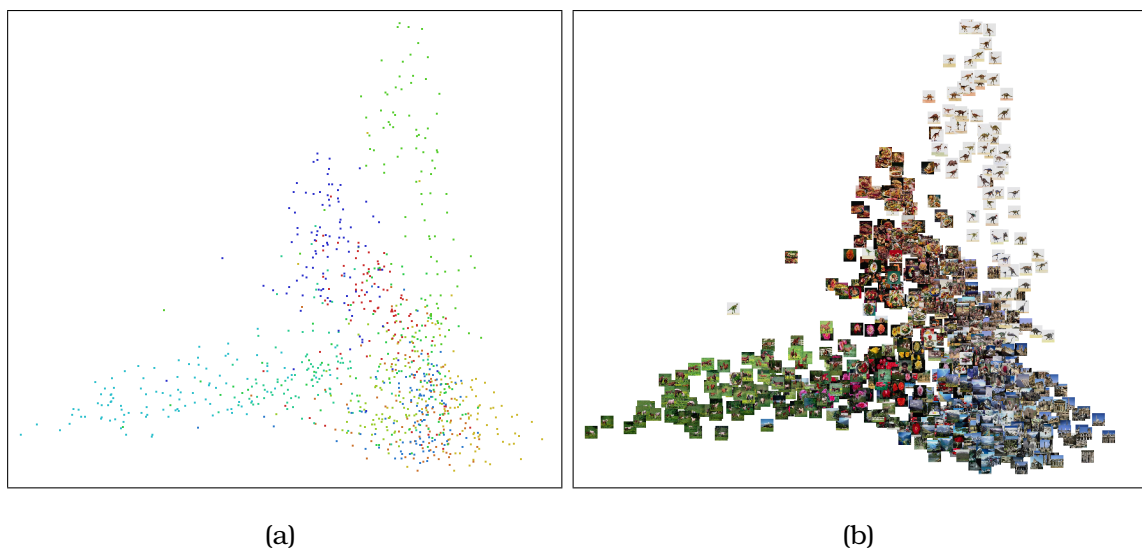


Figura 2.1: Um exemplo de projeção 2D do conjunto de imagens *Corel* (Li and Wang (2003)). O conjunto possui 10 classes de cenas naturais com 100 imagens por classe. Em (a) os elementos visuais são representados por pontos e em (b) a própria imagem é utilizada como rótulo.

## 2.2 Técnicas de Projeção

Existem diversas técnicas que podem ser empregadas para projeção e, cada vez mais, extensões e combinações entre tais técnicas têm sido publicadas (Chen (2006); Paulovich et al. (2008)). Nesta seção, são apresentadas as técnicas mais conhecidas para análise visual de dados. Além disso, as taxonomias existentes na literatura variam de acordo com a função injetiva empregada (Keim et al. (2008); Paulovich (2008)). A taxonomia adotada nesta dissertação divide as técnicas em dois grupos: *lineares* e *não-lineares*. As técnicas de projeção linear são caracterizadas por criar combinações lineares dos atributos, definidas em uma nova base ortogonal de menor dimensão. Em contraste com o grupo de técnicas lineares, as técnicas não-lineares buscam minimizar a função de perda de informação da projeção. Além disso, nesta dissertação é descrita outra técnica de visualização, chamada de *Neighbor-Joining* (Seção 2.2.5) que utiliza princípios de *filogenia* para mapear os dados.

## 2.2.1 Sammon's Mapping

*Sammon's Mapping* - SM é uma técnica de projeção *não-linear* que representa as instâncias de dados em um espaço bidimensional (Sammon (1969)). Inicialmente a matriz de distâncias  $d_{ij}$  é obtida para cada par de entrada  $\{i, j\}$  do conjunto de dados. Para cada instância do espaço original é associada sua imagem no espaço 2D. Para isso, uma nova matriz é computada contendo as distâncias pareadas das imagens calculadas pela norma Euclidiana do vetor, em que  $r_i$  e  $r_j$  são as posições  $x_i$  e  $x_j$  da visualização, respectivamente. O objetivo da técnica é minimizar a função erro  $E$ , ou função de *stress*, definida por:

$$E = \frac{1}{\sum_i \sum_{j>i} d_{ij}} \sum_i \sum_{j>i} \frac{(d_{ij} - \|r_i - r_j\|)^2}{d_{ij}} \quad (2.1)$$

Esta função calcula a soma das diferenças das distâncias obtidas para cada par de pontos. A limitação da técnica de Sammon é a complexidade  $O(n^3)$ , em que  $n$  corresponde o número de instâncias. Martin-Merino and Muñoz (2004) propuseram uma extensão para tratar de visualizações com grande quantidade de dados.

## 2.2.2 Self-Organizing Map

*Self-Organizing Map* - SOM, ou Mapa de Kohonen, é um tipo de rede neural baseada em aprendizagem não supervisionada para gerar uma representação visual bidimensional em forma de grade (Kohonen (1981)). Inicialmente, os nós (neurônios) são localizados sobre a grade com estrutura retangular ou hexagonal. A ideia é que instâncias similares produzam respostas similares em uma rede treinada, e, conseqüentemente, elas sejam agrupadas proxima-mente na representação final (Kohonen et al. (2001)).

Similarmente a uma rede neural, os pesos são ajustados de acordo com as instâncias mais representativas do conjunto. Cada neurônio  $k$  é representado por um vetor  $x_k = (x_{k1}, x_{k2}, \dots, x_{km})$ , tal que  $m$  é a dimensão do espaço de características. Para cada passo do treinamento, um elemento  $d_{ij}$  é selecionado, e a unidade mais próxima é encontrada no mapa. Esta unidade é definida por Kohonen como BMU - *best matching unit*. Em seguida, os vetores da BMU e seus vizinhos na grade são movidos para o vetor referência:

$$x_{(k+1)} = x_k + \alpha^{(t)} h_{ck}^{(t)} (d_{ij} - x_k) \quad (2.2)$$

em que  $\alpha^{(t)} \in [0, 1]$  é a taxa de aprendizagem em iterações  $t$  e  $h_{ck}^{(t)}$  é função de vizinhança centrada sobre a unidade vencedora  $c$ . Tipicamente, a função de vizinhança é  $h_{ck} = \exp(-d^2(c,k)/2\sigma^2)$ , cujo  $\sigma$  especifica o desvio padrão da vizi-

nhança. Embora esta técnica seja útil em muitas aplicações, não há regras para selecionar os melhores parâmetros  $\alpha^{(t)}$  e  $h^{(t)}$ , além de não garantir sua convergência. Outra limitação desta técnica é o custo computacional quadrático associado ao tempo de atualização do mapa.

### 2.2.3 *Interactive Document Map*

A técnica *Interactive Document Map* - IDMAP foi particularmente desenvolvida para a análise de coleções de documentos (Minghim et al. (2006)). Inicialmente, uma projeção das instâncias é criada por meio da técnica *FastMap* (Faloutsos and Lin (1995)), seguido pelo refinamento da *ForceScheme* (Tejada et al. (2003)). A técnica *ForceScheme* é usada para obter a representação final, preservando, o mais fielmente possível, as relações de distância. O objetivo da combinação das duas técnicas é reduzir o alto custo computacional da *ForceScheme* mediante a geração de um *layout* inicial com a técnica *FastMap*, cujo custo é  $O(n)$ . Embora a técnica gere representações que preservam relações de similaridade, ela apresenta limitações de tempo quando utilizada em grandes coleções de documentos. Uma nova extensão foi criada para reduzir o custo quadrático, definida como *Projection by Clustering* - ProjClus (Paulovich and Minghim (2006)).

### 2.2.4 *Least Square Projection*

Uma técnica de projeção *não-linear* e que produz resultados satisfatórios, tanto para aspectos gráficos quanto para preservar relações de distância, é a *Least Square Projection* - LSP (Paulovich et al. (2008)). O seu primeiro passo é eleger os pontos de controle de um subconjunto das instâncias de dados extraídos do conjunto original, denominados pontos de controle. Estes são então projetados no plano por qualquer método de redução de dimensionalidade. O segundo passo utiliza as coordenadas cartesianas dos pontos de controle e as relações de similaridade entre objetos do espaço original para formar um sistema linear que define um novo conjunto de coordenadas cartesianas de posicionamento. Esta técnica atende dois pontos importantes ao gerar projeções: escalabilidade para valores até 100 mil pontos e alta precisão no posicionamento. Entretanto, a sua complexidade é da ordem de  $O(P + S)$ , em que  $P$  corresponde à complexidade de escolher os pontos de controle com custo  $O(n\sqrt{n})$  e  $S$  é a complexidade para resolver o sistema linear com custo  $O(n\sqrt{k})$ . Os índices  $n$  e  $k$  representam o número de pontos de controle e  $k$  a ordem do sistema. Assim, o custo final da LSP é  $O(\max\{n, \sqrt{n}, n\sqrt{k}\})$ .



## 2.2.5 Neighbor-Joining

*Neighbor-Joining* ou *Neighbor-Joint* - NJ é uma técnica *filogenética* que constrói uma árvore com topologia criada a partir da matriz de distâncias, segundo algum critério de reconstrução (Saitou and Nei (1987)). O critério de reconstrução da árvore busca encontrar os pares de instâncias com menor valor de distância e os elegem vizinhos. Dessa maneira, este critério é também conhecido como critério de evolução mínima, sendo obtido pelo tamanho total das ramificações:

$$S_{ij} = \frac{1}{2(n-2)} \left[ \sum_{k \neq i, j} (d_{ik} + d_{jk}) + \frac{1}{2}d_{ij} + \frac{1}{n-2} \sum_{(k, l \neq i, j) \& (k < l)} d_{kl} \right] \quad (2.3)$$

onde  $n$  é o número de vizinhos (no mínimo 2 e no máximo  $n/2$ ),  $d_{ij}$  é o valor na matriz de distâncias e os índices  $k$  e  $l$  representam todos objetos, exceto os selecionados  $i$  e  $j$ . Além disso, o critério de reconstrução é responsável tanto pelo tamanho das ramificações em cada estágio do agrupamento de vizinhos, quanto pela criação das bifurcações presentes na topologia da árvore. A principal limitação desta técnica é o seu custo  $O(n^3)$ . Nos casos em que o número de vizinhos é muito alto, o critério avalia uma pequena porção de todas as topologias possíveis. Embora a técnica não necessariamente produza uma árvore mínima evolutiva, ela garante a topologia correta da árvore (Saitou and Nei (1987)). A Figura 2.2 mostra uma árvore construída pelo algoritmo *Neighbor-Joining*, representada por círculos (Figura 2.2(a)) e pelas próprias imagens (Figura 2.2(b)). As cores dos círculos indicam a classe (categoria) das instâncias de dados.

## 2.2.6 Multidimensional Scaling

Estudos sobre *Multidimensional Scaling* - MDS tiveram origem em psicometria, em que pesquisadores estavam interessados em atribuir valores à observação de dados segundo alguma regra (Seung and Lee (2000); Zhang et al. (2010)). A MDS é uma família de técnicas que representa medidas de similaridade (ou dissimilaridade) entre pares de objetos como distâncias entre pontos de um espaço de baixa dimensão (Cox and Cox (2001)). Tipicamente, a classe de técnicas MDS busca aproximar as distâncias entre pontos  $d_{ij}$  de suas dissimilaridades  $\delta_{ij}$  para todos os pares de objetos  $\{i, j\}$  do espaço original, de tal maneira que:

$$d_{ij} \approx f(\delta_{ij}) \quad (2.4)$$

em que  $f$  é uma função paramétrica monotônica contínua que tenta transfor-

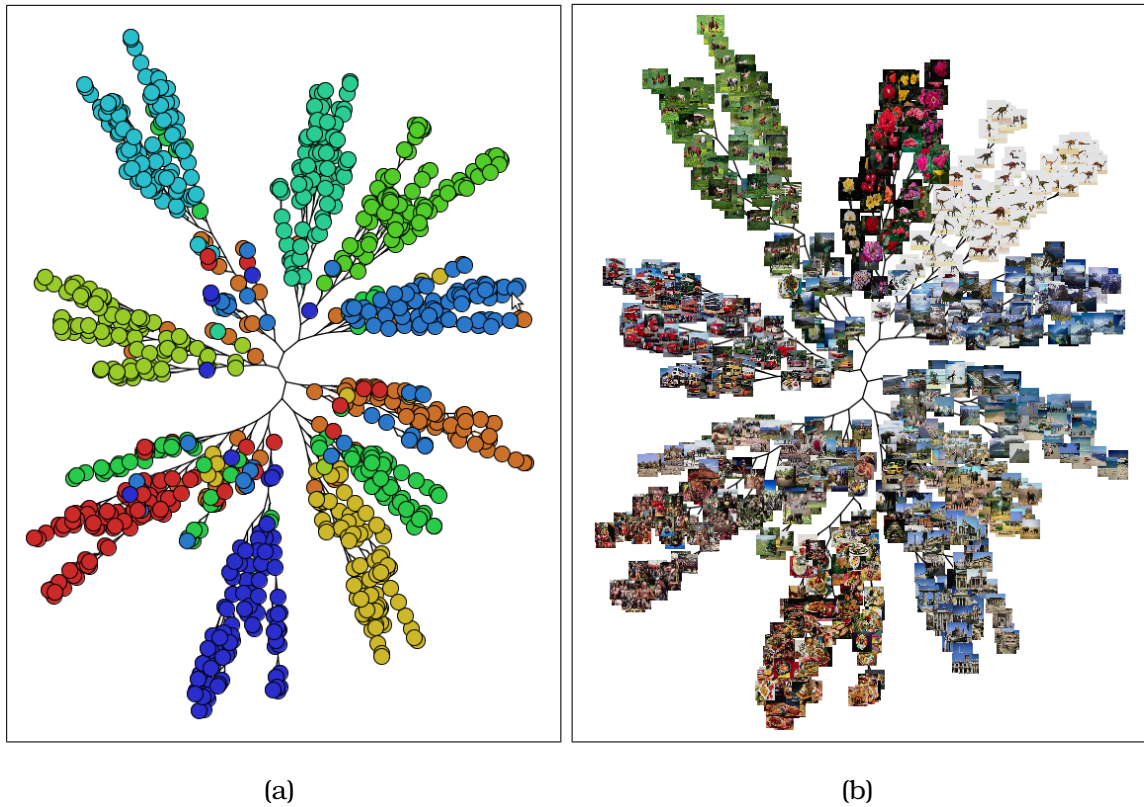


Figura 2.2: Um exemplo de projeção das imagens do conjunto *Corel* com a técnica *Neighbor-Joining*. O vetor de característica é composto por características extraídas de histogramas de cor. Em (a) os elementos visuais são representados por círculos e em (b) a própria imagem é utilizada como rótulo.

mar a medida de proximidade em distâncias (Borg and Groenen (2005); Zhang et al. (2010)).

Matematicamente, suponha as  $n$  instâncias do conjunto  $O$  com dissimilaridades  $\delta_{ij}$  entre os objetos  $\{i, j \in O\}$ , definidas sobre  $O \times O$ . Dado  $\phi$ , um mapeamento arbitrário de  $O$  para  $E$ , em que  $E$  é geralmente um espaço Euclidiano, em que um conjunto de pontos são usados para representar os objetos. Assim, obtém-se que  $\phi(i) = x_i (i \in O, x_i \in E)$  e,  $X = \{x_i \in O\}$  é o conjunto imagem. A distância entre os pontos  $x_i$  e  $x_j$  em  $X$  é dado por  $d_{ij}$ . O objetivo é encontrar um mapeamento  $\phi$ , para que  $d_{ij}$  é aproximadamente igual a  $f(\delta_{ij})$  para todo  $i, j \in O$ .

De acordo com Cox and Cox (2001), a razão por existir diferentes modelos de MDS é devido a diferentes noções da aproximação definida na Equação 2.4, em muitos casos, causada por transformações nas medidas de proximidade. A próxima seção descreve uma técnica métrica do modelo MDS, a *Classical Scaling*. O sucesso pelo uso desta técnica em diversas áreas de conhecimento, é atribuído pela seguinte proposição: “Dado um espaço métrico multidimensional sobre um conjunto de instâncias de dados, existe um espaço Euclidiano em baixa dimensão, cujo a *Classical Scaling* garante encontrar uma estrutura que

preserve tal espaço métrico” (Cox and Cox (2001); de Silva and Tenenbaum (2003); Borg and Groenen (2005)).

### 2.2.6.1 Classical Scaling

A *Classical Multidimensional Scaling* foi a primeira técnica proposta da família MDS, por isso seu nome é conhecido como *Classical Scaling* - CS ou também referido como técnica de Torgerson (Torgerson (1952)). A ideia básica da *Classical Scaling* é assumir que as dissimilaridades são distâncias e depois encontrar as coordenadas do espaço. O método para encontrar as coordenadas Euclidianas originais das distâncias foi proposto por Eckart and Young (1936) e Young and Householder (1938).

Um conjunto de coordenadas de  $n$  pontos em um espaço Euclidiano  $p$ -dimensional é dado por  $\mathbf{x}_i$  para  $(i = 1, \dots, n)$ , em que  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ . Então a distância Euclidiana entre o  $i$ -ésimo e  $j$ -ésimo pontos é dado por:

$$D^2 = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \quad (2.5)$$

Com a matriz de coordenadas  $X$ , é possível calcular o quadrado das distâncias Euclidianas como produtos escalares, produzindo uma nova matriz  $B$ . A matriz de produtos escalares  $B$  (ou matriz de produto interno) é calculada por:

$$B = \mathbf{x}_i^T \mathbf{x}_j \quad (2.6)$$

Entretanto, ambas as coordenadas e a matriz quadrada de distâncias são desconhecidas. Assume-se então que a configuração dos pontos é colocada na origem. Assim:

$$\sum_{i=1}^n x_{ik} = 0 \text{ para } (k = 1, \dots, p) \quad (2.7)$$

Sabe-se então que:

$$D^2 = \mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j - 2\mathbf{x}_i^T \mathbf{x}_j \quad (2.8)$$

onde  $a_{ij} = -\frac{1}{2}D^2$ , e

$$a_{i.} = n^{-1} \sum_j a_{ij}; a_{.j} = n^{-1} \sum_i a_{ij}; a_{..} = n^{-2} \sum_i \sum_j a_{ij}. \quad (2.9)$$

em que  $.$  representa um outro parâmetro de entrada. A matriz  $A$  define-se como  $A_{ij} = a_{ij}$  e a matriz de produtos internos  $B$ :

$$B = HAH \quad (2.10)$$

A matriz  $A$  é multiplicada por ambos os lados, esquerdo e direito, pela matriz de centragem  $H$  definida como:

$$H = I - n^{-1}11^T \quad (2.11)$$

com  $1 = (1, 1, \dots, 1)^T$ , um vetor coluna de uns.

Para encontrar  $B$ , é necessário aplicar a matriz de centragem da Equação 2.11, com fator de  $-\frac{1}{2}$  encontrado em 2.9, multiplicado pela matriz de distâncias ao quadrado  $D^2$ :

$$-\frac{1}{2}HD^2H = B \quad (2.12)$$

A operação da Equação 2.12 é chamada de centragem dupla. Para recalcular as coordenadas de  $B$ , a matriz de produtos internos pode ser escrita como:

$$B = XX^T \quad (2.13)$$

em que  $X = [\mathbf{x}_1, \dots, \mathbf{x}_n]^T$  representa a matriz de coordenadas  $n \times p$ . O *rank* de  $B$ ,  $r(B)$  é então:  $r(B) = r(XX^T) = r(X) = p$ . Isto significa que  $B$  é uma matriz simétrica, positiva e de *rank*  $p$ , tal que existem  $p$  autovalores não negativos. Dessa maneira, para encontrar as coordenadas de  $B$ , fatora-se  $B$  e computa-se sua decomposição espectral:

$$B = Q\Lambda Q^T \quad (2.14)$$

em que  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , a matriz diagonal de autovalores  $\{\lambda_i\}$  de  $B$ , e  $Q = [v_1, \dots, v_n]$ , a matriz de autovetores correspondente, normalizados para 1. Como em Classical Scaling, os autovalores negativos são ignorados como erro, a Equação 2.16 pode ser reescrita:

$$B = Q_+\Lambda_+Q_+^T \quad (2.15)$$

A solução de  $B = XX^T$  é calculado pela matriz de coordenada  $X$  dado por:

$$X = Q_+\Lambda_+^{\frac{1}{2}} \quad (2.16)$$

onde  $\Lambda_+^{\frac{1}{2}} = \text{diag}(\lambda_+^{\frac{1}{2}}, \dots, \lambda_p^{\frac{1}{2}})$  para as coordenadas dos pontos recalculados das distâncias entre os pontos.

A técnica *Classical Scaling* somente difere-se deste procedimento substituindo a matriz de distâncias quadrada  $D^2$  pelo quadrado da matriz de dissimilaridades  $\Delta^2$ . Os passos para a Classical Scaling são mostrados no algo-

ritmo 1.

---

**Algoritmo 1:** Algoritmo da técnica *Classical Scaling*.

---

**Entrada:** Conjunto de observações, instâncias de dados.

**Saída:** Matriz de coordenadas.

- 1 Calcula a matriz de dissimilaridade ao quadrado  $\Delta^2$ .
  - 2 Aplica centragem dupla sobre matriz do passo 1, tal que  $B_\Delta = -\frac{1}{2}H\Delta^2H$ .
  - 3 Computa a decomposição espectral de  $B_\Delta = Q\Lambda Q^T$ .
  - 4 Seja a matriz dos primeiros autovalores maiores que zero  $\Lambda_+$  e  $Q_+$  a primeira coluna de  $Q$ . Assim, a matriz de coordenadas da *Classical Scaling* é dada por  $X = Q_+\Lambda_+^{\frac{1}{2}}$ .
- 

Embora a redução de dimensionalidade com a técnica *Classical Scaling* preserve o espaço métrico, o seu custo computacional é da ordem de  $O(n^3)$  em que  $n$  representa o número de instâncias. Este alto custo está associado ao cálculo dos autovalores e autovetores da matriz  $N \times N$ , derivada da matriz de distância das instâncias de entrada. Uma variante da MDS que preserva as propriedades da *Classical Scaling* foi proposto por de Silva and Tenenbaum (2003). Em contraste com a *Classical Scaling*, a técnica *Landmark MDS - LMDS* escolhe um subconjunto dos dados, referido como *landmarks points*, em um espaço de baixa dimensão. Não somente a complexidade cai para custo linear, mas a técnica também consegue descrever uma boa aproximação para o algoritmo da *Classical Scaling* mesmo com a presença de ruído, porém é necessário uma maior quantidade de *landmarks*.

### 2.3 Avaliação da Qualidade da Projeção

O objetivo em avaliar a qualidade da representação visual é reduzir a subjetividade durante a análise de diferentes projeções. Embora a habilidade de usuários possa ser empregada para determinar a qualidade de uma projeção, muitas vezes é difícil apontar diferenças na formação de grupos e subgrupos quando há mudanças apenas nas características. Assim, uma projeção pode ser analisada como um agrupamento, tornando-se bastante próxima a relação com validação de agrupamentos (Kaufman and Rousseeuw (1990)). Em agrupamento de dados, a validação de agrupamento visa avaliar os resultados de um algoritmo de agrupamento (Everitt and Leese (2001)). Esta validação consiste em três critérios: *externo*, *interno* e *relativo*. Para a avaliação de projeções, a qualidade pode ser medida tanto por meio de critérios de validação de agrupamentos, quanto por medidas estatísticas.

Do ponto de vista de agrupamentos, as estratégias mais indicadas para as projeções são os critérios *interno* e *relativo*, pois muitas vezes não se pode definir o agrupamento de referência requerido por qualquer índice do critério

*externo*<sup>1</sup>. Em estudos comparativos, Vendramin et al. (2009) estendeu o trabalho de Milligan and Cooper (1985), e concluiu que o *índice da silhueta* foi o melhor avaliado em relação a 30 índices internos. Este critério será descrito na Seção 2.3.1. Vale ressaltar que existem trabalhos que definem o *índice da silhueta* como índice *relativo*, pois índices *internos* podem ser empregados em critério *relativo* (Halkidi et al. (2002)).

Além dos conhecidos critérios para validação de agrupamentos, existem medidas estatísticas para avaliar a qualidade de projeções sobre um conjunto de pontos (nuvem de pontos) no espaço Euclidiano (Wilkinson et al. (2006)). Estas medidas incluem: *densidade*, *assimetria*, *correlação*, *forma* e *outliers*. A medida *projection precision score - pps*, proposta por Schreck et al. (2010), compara a distância de um ponto e seus vizinhos entre o espaço original e o espaço projetado. Estas são medidas encontradas em trabalhos de visualizações de estrutura de dados. Porém, uma vasta literatura é dedicada à análise de agrupamentos (Milligan and Cooper (1985); Xu and Wunsch (2005); Vendramin et al. (2009)).

### 2.3.1 Índice da Silhueta

O *índice da silhueta*, ou *coeficiente de silhueta*, avalia a qualidade de um agrupamento com as diferenças entre as distâncias intra-grupo e inter-grupo das instâncias (Rousseeuw (1987)). Considere uma instância  $i$  pertencente a um grupo. Sua coesão  $a_i$  é calculada pela média das distâncias entre  $i$  e todas as outras instâncias pertencentes ao mesmo grupo (*relação intra-grupo*). A separação  $b_i$  é a distância mínima entre  $i$  e todas outras instâncias de outros grupos (*relação inter-grupo*). O índice da silhueta  $S$  de um agrupamento é calculado pela média da silhueta de todas as instâncias  $n$ , definido por:

$$S = \frac{1}{n} \sum_{i=1}^n \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (2.17)$$

onde o denominador é apenas um termo para normalização. O índice da silhueta pode variar entre  $-1 \leq S \leq 1$ . Se o agrupamento possui apenas um elemento, considerado um *singleton*, assume-se que o valor da silhueta é zero. Os valores mais altos para  $S$  indicam melhor coesão e separação entre os grupos, que significa minimizar a distância intra-grupo e maximizar a distância inter-grupo. O *índice da silhueta* é caracterizado por privilegiar partições com grupos compactos e bem separados (Kaufman and Rousseeuw (1990)). Claramente, o custo para o cálculo da silhueta é  $O(n^2)$ .

---

<sup>1</sup>Um critério de validação é uma estratégia para avaliar agrupamentos, enquanto um índice é uma medida objetiva adotando tal estratégia.

## 2.4 Ferramenta PEx-Image

*Projection Explorer for Images* ou *PEx-Image*<sup>2</sup> é uma ferramenta desenvolvida em linguagem Java para usuários explorarem visualmente conjuntos de imagens em um espaço bidimensional (Eler et al. (2008, 2009)). O principal objetivo é permitir que usuários possam explorar diferentes espaços de características ao modificar algoritmos ou parâmetros utilizados para a descrição de imagens. Além disso, a ferramenta proporciona o uso da sua interface como recurso *front-end* para técnicas criadas e implementadas pelo *Grupo de Visualização, Imagens e Computação Gráfica - VICG*, no *Instituto de Ciências Matemáticas e de Computação - ICMC, Universidade de São Paulo - USP*. Vale ressaltar que a ferramenta *PEx-Image* é uma extensão da ferramenta *Projection Explorer - PEx* (Paulovich et al. (2007)), cujo propósito é explorar coleções de documentos (Paulovich et al. (2006); Paulovich and Minghim (2006)). Ambas ferramentas possuem recursos de interação que fornecem ao usuário meios para focar em uma determinada região e atingir maiores níveis de detalhes. Estes recursos são também conhecidos na literatura como *técnicas de interação* (Keim (2000)).

Conforme mencionado na Seção 2.1, cada instância do conjunto de imagens é representada como um elemento visual na projeção, como um círculo, ponto, ou a própria imagem. A Figura 2.3 mostra uma tela capturada da interface da ferramenta *PEx-Image* explorando o conjunto de imagens *Corel*. Neste caso específico, o objetivo é auxiliar o usuário na descoberta de conhecimento, na criação de novas hipóteses e no entendimento do comportamento dos descritores que falham na separabilidade dos grupos *pratos* e *flores*, incitando-o a explorar uma outra característica. Dessa maneira, as técnicas de projeção devem preservar o mais fielmente possível as relações de similaridade. Além disso, as medidas para a avaliação das projeções ajudam a reduzir a subjetividade no processo de exploração.

## 2.5 Considerações Finais

Definir quais técnicas de visualização podem ser empregadas em processos de *mineração visual de dados* não é uma tarefa trivial, pois cada técnica possui um conjunto bastante particular de características (Fayyad et al. (2001)). Algumas delas são capazes de manipular diferentes tipos de entradas de dados, enquanto outras são melhores em identificar grupos e correlações presentes no conjunto de dados. Além disso, o custo computacional das técnicas deve

---

<sup>2</sup>Disponível para download em <http://infoserver.lcad.icmc.usp.br/infovis2/PExImage>

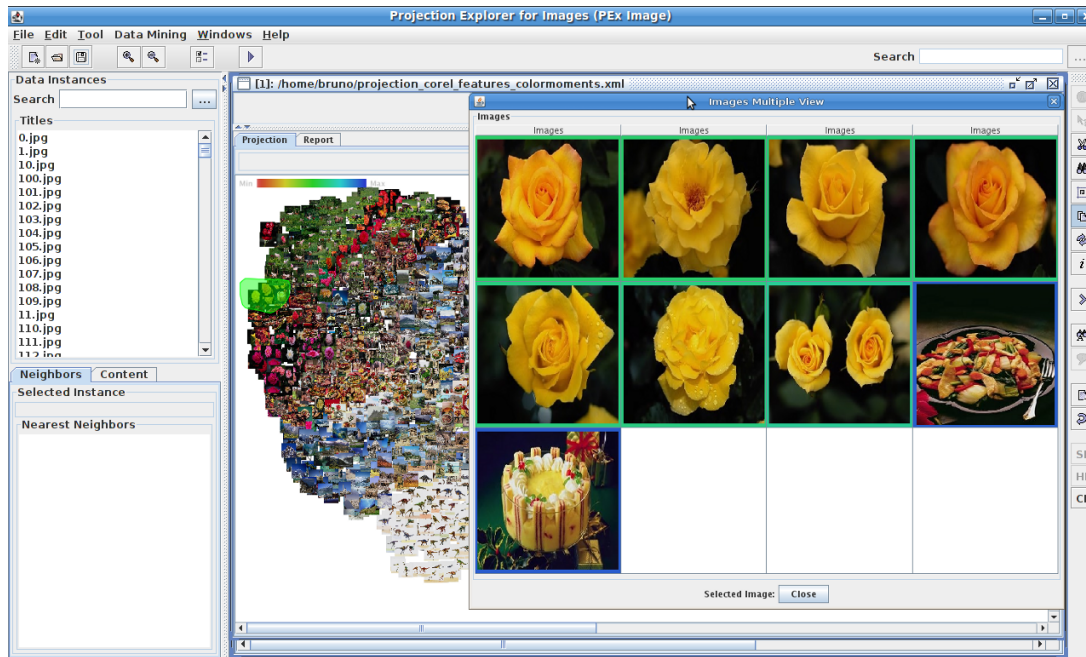


Figura 2.3: Um exemplo de exploração visual do espaço de características com a ferramenta *PEX-Image*. Neste caso específico, o usuário usa recursos de interação para explorar as imagens dos grupos *pratos* e *flores* do conjunto *Corel*.

ser levado em consideração, principalmente para grandes conjuntos de dados. A Tabela 2.1, ordenada cronologicamente, mostra a complexidade das técnicas de projeção discutidas neste capítulo. Os índices  $n$  e  $p$  representam o número de instâncias e amostras respectivamente,  $k$  corresponde ao número de dimensões do espaço, e  $c$  é uma constante de iteração.

Tabela 2.1: Principais técnicas de projeção, seus respectivos autores e custos computacionais.

<b>Técnica de Projeção</b>	<b>Autor</b>	<b>Complexidade</b>
<i>Classical Scaling</i>	Torgerson (1952)	$O(n^3)$
<i>Sammon's Map</i>	Sammon (1969)	$O(n^3)$
<i>SOM</i>	Kohonen (1981)	$O(cn^2)$
<i>Neighbor-Joining</i>	Saitou and Nei (1987)	$O(n^3)$
<i>FastMap</i>	Faloutsos and Lin (1995)	$O(n)$
<i>LMDS</i>	de Silva and Tenenbaum (2003)	$O(k^3 + kn)$
<i>ForceScheme</i>	Tejada et al. (2003)	$O(cn^2)$
<i>IDMAP</i>	Minghim et al. (2006)	$O(cn^2)$
<i>LSP</i>	Paulovich et al. (2008)	$O(\max\{n, \sqrt{n}, n\sqrt{k}\})$

A maior crítica em relação às técnicas de projeção é a inevitável perda de informação quando os dados são projetados para espaços 2D ou 3D. Essa crítica pode ser reduzida pela escolha cuidadosa das técnicas de projeção que mais preservam as relações de similaridade entre as instâncias. Matematicamente, a técnica *Multidimensional Classical Scaling* é uma das técnicas que



mais refletem tais relações entre o espaço original e a representação final (Cox and Cox (2001)).

Os critérios *interno* e *relativo* são as duas estratégias de agrupamento que podem ser empregadas para validar objetivamente as projeções. Este tipo de critério é determinado por utilizar as relações de similaridade do próprio conjunto, ao contrário do critério *externo* que utiliza uma partição de referência. Na Seção 2.3.1, descreveu-se o *índice da silhueta* para avaliar a qualidade das projeções.

Por fim, na Seção 2.4 foi apresentada a ferramenta *PEX-Image* desenvolvida para explorar visualmente conjuntos de imagens (Eler et al. (2009)). Apesar de existir um número considerável de grupos empenhados em desenvolver ferramentas para *mineração visual de dados* (Swayne et al. (2003); Li (2004); Chen (2006)), poucos deles exploram conjuntos de imagens por meio de projeção. Além disso, existem propostas de ferramentas similares ao *PEX-Image*, porém elas se dedicam a organizar painéis de projeções 2D e guiar a exploração mediante a *rankings*, criados de acordo com medidas selecionadas pelo usuário (Seo and Shneiderman (2005, 2006); Wilkinson et al. (2006)).



---

## DESCRIÇÃO DE IMAGENS

---

Como apresentado no capítulo anterior, as representações visuais podem auxiliar especialistas a identificar padrões em conjuntos de dados multidimensionais, tais como imagens. Entretanto, a similaridade das instâncias mapeadas na projeção é definida segundo uma função de distância e um conjunto de características representado em um vetor.

Este capítulo apresenta técnicas e estratégias para computar vetores de características de imagens. A Seção 3.1 introduz a motivação em representar uma imagem por meio de características. Uma divisão sobre os grupos de técnicas é apresentada na Seção 3.2. Em seguida, são descritos dois grupos mediante suas propriedades físicas. O primeiro deles apresenta técnicas baseadas em *textura* (Seção 3.3) e o segundo grupo descreve técnicas baseadas na informação de *cor* (Seção 3.4). Na Seção 3.5 são discutidas a representação de imagens usando *características locais* e o modelo *bag-of-visual-features*. Por fim, as considerações finais do capítulo são feitas na Seção 3.6.

### 3.1 Considerações Iniciais

Uma importante etapa em sistemas de análise de imagens presentes em *visão computacional*, *recuperação baseada em conteúdo* e *reconhecimento de padrão*, é a descrição de imagens. O objetivo das técnicas de descrição, ou simplesmente descritores, é extrair medidas quantitativas de um conjunto de imagens para classificação de objetos. Tipicamente, um descritor reduz a dimensionalidade das imagens de  $\mathbb{R}^{w \times h}$  para  $\mathbb{R}^n$ , em que  $n$  é a quantidade de características extraídas,  $w$  e  $h$  são a largura e a altura da imagem, respectivamente. Além disso, o *processamento de imagens* pode ser empregado pri-

mariamente para realçar o conteúdo da imagem e para segmentar os objetos de interesse dentro da cena.

A multidisciplinaridade desta subárea de pesquisa nos conduz a encontrar diferentes terminologias na literatura. Grande parte dos pesquisadores que trabalham com *análise semântica*, definem descrição como extração de características visuais (Datta et al. (2008)), pois existem outras medidas que podem ser computadas para classificação de objetos, por exemplo, características de áudio para sistemas de reconhecimento de voz. Por um outro ponto de vista, Sonka et al. (2008) definem extração de características como a combinação das etapas de descrição de imagens e seleção de características. Para esta dissertação, o termo extração de características refere-se apenas à etapa de descrição.

### 3.2 Técnicas para Descrição de Imagens

Existem diversas técnicas, e um grande número de extensões, criadas para descrição de imagens (Datta et al. (2008)). Primeiramente, é necessário considerar que uma imagem pode ser representada ou em termos de suas características externas (*contorno*), ou em termos de suas características internas (*região*). A representação externa é escolhida quando as imagens podem ser descritas pela *forma* dos objetos, enquanto a representação interna foca nas propriedades dos pixels contidos em uma região, como *cor* e *textura*. Definido um esquema de representação, a próxima tarefa é empregar uma técnica de descrição. O presente trabalho emprega a representação interna, com foco em textura e cor, devido tais informações serem mais evidentes nos conjuntos de imagens. Neste capítulo, são apresentadas as técnicas baseadas em *textura* e *cor*.

### 3.3 Análise de Textura

O estudo sobre a percepção e discriminação de texturas teve origem com os trabalhos de Julesz (1962, 1966), em *psicofísica*. Atualmente, a análise de textura continua sendo uma importante área de pesquisa devido à vasta gama de aplicações, tais como identificação de tumores em *imagens médicas* e inspeção automática em *sensoriamento remoto* (Cheng et al. (2010)). Tuceryan and Jain (1998) criaram uma taxonomia que divide as técnicas de análise de textura em quatro grandes grupos: *estruturais*, *estatísticas*, *espectrais* e *baseadas em modelos*. As próximas seções descrevem duas técnicas tradicionais em análise de textura, matriz de co-ocorrência e filtros de Gabor. A primeira pertence ao grupo estatístico, enquanto a segunda é uma técnica espectral.

### 3.3.1 Matriz de Co-ocorrência

Matriz de co-ocorrência é uma técnica *estatística de segunda ordem* que descreve a textura de acordo com a co-ocorrência dos *níveis de cinza* da imagem. A técnica considera a posição relativa dos pixels e quantifica padrões presentes em texturas, tais como rugosidade e direcionalidade. As matrizes  $M_{d\theta}$  são computadas para diferentes parâmetros de distância  $d$  e orientação  $\theta$  da imagem  $I$ , segundo o número de ocorrência que os pares de pixels com intensidades  $g_i, g_j$  aparecem em  $I$ , tal que  $1 < i, j < L$  correspondem aos níveis de cinza (intensidade). A Equação 3.1 gera as matrizes de co-ocorrência.

$$M_{d\theta}(i, j) = \sum_{x=1}^w \sum_{y=1}^h \begin{cases} 1, & \text{se } I(x, y) = i \text{ e } I(x', y') = j \\ 0, & \text{caso contrário} \end{cases} \quad (3.1)$$

onde  $I(x, y)$  é o nível de cinza na posição  $(x, y)$ ,  $x' = x + d \cos \theta$  e  $y' = y + d \sin \theta$ .

A matriz  $M_{d\theta}$  resultante é da ordem de  $G^2$ , tal que  $G$  é o número de *níveis de cinza* da imagem. O vetor de característica é composto por medidas estatísticas calculadas da matriz. Haralick et al. (1973) propuseram 14 medidas estatísticas, porém poucas delas são normalmente utilizadas devido à redundância. Em diversos trabalhos (Drimbarean and Whelan (2001)), e para esta dissertação, são utilizadas cinco medidas descritas na Tabela 3.1, que são *energia, entropia, inércia, momento de diferença inversa e correlação*.

Tabela 3.1: Medidas estatísticas calculadas por meio das matrizes de co-ocorrências. A primeira coluna da tabela contém o nome das medidas e a segunda coluna apresenta a formalização da referida medida.

	Medida Estatística	Formalização
(E)	<i>Energia</i>	$\sum_{i,j} M_{d\theta}(i, j)^2$
(H)	<i>Entropia</i>	$-\sum_{i,j} M_{d\theta}(i, j) \log M_{d\theta}(i, j)$
(I)	<i>Inércia</i>	$\sum_{i,j} M_{d\theta}(i, j)(i - j)^2$
(IDM)	<i>Momento de Diferença Inversa</i>	$\sum_{i,j} \frac{1}{1 + (i - j)^2} M_{d\theta}(i, j)$
(COR)	<i>Correlação</i>	$\frac{\sum_{i,j} (i, j) M_{d\theta}(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$

Os parâmetros  $\mu_x$ ,  $\mu_y$ ,  $\sigma_x$  e  $\sigma_y$  correspondem às médias e aos desvios pa-

drão de  $M_{d\theta_x}$  e  $M_{d\theta_y}$ . Por fim, o vetor de característica  $\psi$  para esta técnica é composto pelas medidas estatísticas apresentadas na Tabela 3.1, calculadas para diferentes distâncias  $d_1, \dots, d_N$  e direções  $\theta_1, \dots, \theta_M$ . Por exemplo, usando 5 medidas, 5 distâncias (de 1 até 5) e 4 direções ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ ), obtém-se um vetor  $\psi$  com 100 características:

$$\psi = [E_{d_1\theta_1}, \dots, COR_{d_1\theta_1}, \dots, E_{d_N\theta_M}, \dots, COR_{d_N\theta_M}] \quad (3.2)$$

### 3.3.2 Filtros de Gabor

Filtros de Gabor é uma técnica *espectral* amplamente utilizada para análise de textura (Bianconi and Fernández (2007)). Uma função Gabor é um sinal senoidal (ou um sinal cosseno) modulado por uma Gaussiana (Gabor (1946)). Os filtros de Gabor utilizados na decomposição de imagens são criados a partir da função Gabor-mãe de duas dimensões, dado domínio do espaço  $g(x, y)$  e frequência  $G(x, y)$ . As Equações 3.3 e 3.4 descrevem a função Gabor-mãe no domínio do espaço e da frequência, respectivamente.

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j W x} \quad (3.3)$$

$$G(x, y) = e^{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]} \quad (3.4)$$

cujos  $\sigma_x$  e  $\sigma_y$  representam o desvio padrão da Gaussiana e correspondem à  $\sigma_u = \frac{1}{2}\pi\sigma_x$  e  $\sigma_v = \frac{1}{2}\pi\sigma_y$  da função Gabor-mãe no domínio da frequência. Dada a função mãe, um banco de filtros de Gabor pode ser obtido no domínio do espaço  $g(x, y)$  por operações de dilatações e rotações, de acordo com a Equação 3.5.

$$g_{mn} = a^{-m}G(x', y') \quad (3.5)$$

para  $a > 1$ ,  $x' = a^{-m}(x \cos \theta + y \sin \theta)$  e  $y' = a^{-m}(-x \cos \theta + y \sin \theta)$ .

Inicialmente, a técnica gera um banco de filtros  $g_{mn}(x, y)$  para diferentes parâmetros escalas  $m = 1, \dots, K$  e orientações  $n = 1, \dots, S$ . Características de textura são computadas pela convolução da imagem original  $I$  com o banco de filtros de Gabor, como descrito na Equação 3.6. Ao variar os valores de  $m$  e  $n$ , alguns aspectos da textura são realçados. Em geral, os trabalhos na literatura utilizam 4 orientações ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ ) e 4 escalas para criar o banco de filtros.

$$c_{mn}(x, y) = I(x, y) * g_{mn}(x, y) \quad (3.6)$$

O vetor de característica  $\psi$  é finalmente obtido ao computar a energia e entropia para as  $mn$ -ésimas imagens filtradas, de acordo com a Equação 3.7.

$$\psi = [E_{11}, E_{12}, \dots, E_{KS}], \text{ dado } E_{mn} = \sum_{x,y} [c_{mn}(x, y)]^2 \quad (3.7)$$

Imagens naturais frequentemente são constituídas de regiões com variação na intensidade dos pixels. Ambas técnicas, matriz de co-ocorrência e filtros de Gabor, foram propostas originalmente para tratar imagens em *níveis de cinza*. Entretanto, há extensões na literatura que testaram ambos descritores em cenas de *textura colorida* (Drimbarean and Whelan (2001)). Atualmente, grande parte dos conjuntos de imagens carregam informação de cor. Na próxima seção são apresentadas técnicas por meio da análise de cor.

### 3.4 Análise de Cor

A cor é uma importante propriedade discriminatória de objetos. Além disso, ela fornece informação extra que permite a distinção entre vários aspectos físicos que causam variações de cor, tais como mudanças de iluminação, reflectância do objeto e da fonte de luz. Inicialmente, na Seção 3.4.1, é apresentada uma taxonomia de propriedades invariantes derivadas do *modelo diagonal* (Forsyth (1990)). Em seguida, são apresentados os descritores de cor mais conhecidos na literatura, os quais estão divididos em três grandes grupos: *histogramas*, *momentos de cor* e descritores baseado em *Scale Invariant Feature Transform* - SIFT.

#### 3.4.1 Modelo de Reflectância em Constância de Cor

A percepção das cores é determinada pela natureza da luz refletida do objeto que varia em função do comprimento de onda no espectro eletromagnético (Gonzalez and Woods (2006)). Entretanto, há diversos outros aspectos que influenciam a formação de cores em imagens, incluindo sombra, reflectância e superfícies dos materiais (van de Weijer (2005)).

Shafer (1985) propôs um modelo físico de reflectância de objetos em *visão computacional* e *computação gráfica*. Este modelo, também conhecido como modelo de reflexão dicromática, identifica como mudanças fotométricas influenciam nos valores dos pixels. O modelo divide a luz refletida em reflexão do corpo do objeto e reflexão de superfície (especular). Este é considerado na literatura um bom modelo físico para generalização do problema de reflectância (Mindru et al. (2004)). Além disso, ele não assume nenhuma fonte de luz ou distribuição de iluminação uniforme sobre a cena. Shafer (1985) estendeu o modelo de reflectância de Lambert para uma imagem  $I$ , adicionando o termo da *luz difusa*  $A(\lambda)$  na Equação 3.8.

$$I(\xi) = \int_{\omega} e(\lambda) s(\xi, \lambda) \rho_k(\lambda) d\lambda + \int_w A(\lambda) \rho_k(\lambda) d\lambda \quad (3.8)$$

onde  $e(\lambda)$  é a cor da fonte de luz,  $s(\xi, \lambda)$  é a superfície de reflectância,  $\rho_k(\lambda)$  é a função de sensibilidade da câmera ( $k \in R, G, B$ ),  $\lambda$  é o comprimento de onda,  $\omega$  e  $\xi$  são o espectro visível e a posição no espaço, respectivamente. A *luz difusa*, causada pela reflexão difusa, ocorre quando um raio de luz atinge a superfície de um material que refletirá de volta para todas as direções em igual intensidade.

O cálculo da derivada da imagem  $I$ , produzirá invariância à *luz difusa*. Anos depois, o *modelo de reflectância* de Lambert foi associado ao *modelo diagonal* de mudança de iluminação criado por Forsyth (1990). Entretanto, Finlayson et al. (2005) mostraram que há um número significativo de imagens naturais em que o *modelo diagonal* falha. Para resolver tal falha, eles propuseram o *modelo diagonal com deslocamento* que será descrito a seguir.

### 3.4.1.1 Modelo Diagonal com Deslocamento

Forsyth (1990) modelou o problema de mudanças de iluminação por meio de uma matriz diagonal com três fatores de escala  $[x, y, z]$  que mapeia as cores de uma imagem  $RGB$  tirada sob uma fonte de luz desconhecida  $u$  para seu espaço  $RGB$  correspondente sob uma luz de referência  $c$  (iluminante canônico) de acordo com:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \quad (3.9)$$

Este modelo é chamado de *modelo diagonal* de mudança de iluminação por causa da matriz diagonal. Apesar da solução ser robusta em grande parte das superfícies do problema de constância de cor, há casos em que o mapeamento não consegue estimar o iluminante da cena. Finlayson et al. (2005) estenderam o *modelo diagonal* adicionando o termo de translação ou deslocamento  $(d_1, d_2, d_3)^T$ , resultando no *modelo diagonal com deslocamento*:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} \quad (3.10)$$

Dessa maneira, assume-se que o *modelo diagonal com deslocamento* é equivalente ao modelo de reflectância com o termo da *luz difusa* proposto por Shafer (1985) (Equação 3.8):



$$\begin{pmatrix} e^c(\lambda_R) \\ e^c(\lambda_G) \\ e^c(\lambda_B) \end{pmatrix} = \begin{pmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{pmatrix} \begin{pmatrix} e^u(\lambda_R) \\ e^u(\lambda_G) \\ e^u(\lambda_B) \end{pmatrix} + \begin{pmatrix} A(\lambda_R) \\ A(\lambda_G) \\ A(\lambda_B) \end{pmatrix} = \begin{pmatrix} x[e(\lambda_R)] + A(\lambda_R) \\ y[e(\lambda_G)] + A(\lambda_G) \\ z[e(\lambda_B)] + A(\lambda_B) \end{pmatrix} \quad (3.11)$$

onde  $\lambda_R$ ,  $\lambda_G$  e  $\lambda_B$  são os comprimentos de onda na posição  $\xi$  com superfície de reflectância  $s(\xi, \lambda_C)$  para  $C \in \{R, G, B\}$ . Esta notação é usada em transformações fotométricas para imagens coloridas.

### 3.4.1.2 Transformações Fotométricas

As transformações fotométricas têm recebido grande atenção devido ao poder discriminatório da informação adicional de cor comparada à informação em *níveis de cinza* (van de Weijer (2005); Burghouts and Geusebroek (2009)). De acordo com o *modelo diagonal com deslocamento*, as transformações fotométricas, e suas correspondentes propriedades, estão divididas em:

1. **Variação na intensidade de iluminação:** inclui sombra e transformações geométricas (Mindru et al. (2004)). Quando uma característica é invariante à esta variação, ela é dita ser invariante à escala em relação à iluminação.
2. **Deslocamento na intensidade de iluminação:** corresponde aos pontos destacados sob uma fonte de luz branca.
3. **Variação e deslocamento na intensidade de iluminação:** é a combinação de ambas as condições (1) e (2).
4. **Variação na cor da fonte de luz:** corresponde à mudança na cor da fonte de luz.
5. **Variação e deslocamento na cor da fonte de luz:** inclui a variação na cor da luz (4) associada ao deslocamento na fonte de iluminação.

## 3.4.2 Histogramas

Um histograma de imagem representa a distribuição dos pixels segundo seus valores de intensidade (Gonzalez and Woods (2006)). Geralmente, os vetores de características de histogramas têm alta dimensionalidade, pois a dimensão está associada ao número de *bins* da distribuição.

### 3.4.2.1 Histograma RGB

O histograma *RGB* é uma combinação de três histogramas unidimensionais dos canais do espaço de cor *RGB* (*red – green – blue*). Histogramas de cor

não consideram informações espaciais locais e sobre conjuntos de imagens com fundos altamente correlacionados, geralmente, o desempenho de classificação é melhorado (Gevers and Stokman (2004)). Este histograma não possui nenhuma propriedade de invariância.

### 3.4.2.2 Histograma Oponente

O histograma oponente também combina três histogramas unidimensionais, porém baseado nos canais do espaço de cor oponente, de acordo com a Equação 3.12.

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(R-G) \\ \frac{1}{\sqrt{6}}(R+G-2B) \\ \frac{1}{\sqrt{3}}(R+G+B) \end{pmatrix} \quad (3.12)$$

$O_3$  representa a intensidade da cor enquanto,  $O_1$  e  $O_2$  representam a informação dos canais de cor, respectivamente. Devido à subtração em  $O_1$  e  $O_2$ , os deslocamentos podem se cancelar, caso eles sejam iguais para todos os canais (Lukac and Plataniotis (2007)). Isto pode ser verificado pela presença de uma fonte de luz branca<sup>1</sup> que causa o deslocamento  $d_1$ :

$$\begin{pmatrix} O_1 \\ O_2 \end{pmatrix} = \begin{pmatrix} \frac{R^c-G^c}{\sqrt{2}} \\ \frac{R^c+G^c-2B^c}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} \frac{(R^u+d_1)-(G^u+d_1)}{\sqrt{2}} \\ \frac{(R^u+d_1)+(G^u+d_1)-(2B^u+d_1)}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} \frac{R^u-G^u}{\sqrt{2}} \\ \frac{R^u+G^u-2B^u}{\sqrt{6}} \end{pmatrix} \quad (3.13)$$

Dessa maneira,  $O_1$  e  $O_2$  não possuem invariância a deslocamento em relação à intensidade de iluminação. O canal  $O_3$  não têm propriedades de invariância em relação às transformações fotométricas apresentadas na Seção 3.4.1.2.

### 3.4.2.3 Histograma Matiz

Este descritor é caracterizado por duas características do espaço de cor *HSI* (*hue – saturation – intensity*). Sabe-se que o modelo *HSI* torna-se instável ao redor do eixo cinza, além de ser inversamente proporcional a sua saturação (van de Weijer et al. (2006)). Assim, este histograma torna-se mais robusto quando cada ponto da matiz é ponderado pela sua saturação. O modelo *HSI* é dado por:

$$\begin{pmatrix} H' \\ S' \\ I' \end{pmatrix} = \begin{pmatrix} \arctan\left(\frac{O_1}{O_2}\right) \\ \sqrt{(O_1)^2 + (O_2)^2} \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \frac{R(B-G)+G(R-B)+B(G-R)}{R^2+G^2+B^2-RG-RB-GB} \\ \frac{1}{\sqrt{6}} \frac{R(2R-G-B)+G(2G-R-B)+B(2B-G-R)}{R^2+G^2+B^2-RG-RB-GB} \\ \frac{1}{\sqrt{3}}(R+G+B) \end{pmatrix} \quad (3.14)$$

<sup>1</sup>Para uma fonte de luz branca assume-se que todos os comprimentos de onda dentro do espectro visível (de 400nm a 700nm) têm energia similar ( $e(\lambda) = e$ ).

em que  $O_3$ , derivado do espaço de cor oponente (Equação 3.12), corresponde à intensidade. Neste caso, o descritor é invariante à escala e ao deslocamento em relação à variação de iluminação.

#### 3.4.2.4 Histograma RG

No modelo de cor  $RGB$  normalizado, as componentes de cromaticidade  $R$  e  $G$  descrevem a informação de cor na imagem, sendo  $B$  redundante conforme  $R + G + B = 1$ :

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \\ \frac{B}{R+G+B} \end{pmatrix} \quad (3.15)$$

Devido à normalização,  $R$  e  $G$  são invariantes à mudança de iluminação e sombras, de acordo com a Equação 3.16:

$$\begin{pmatrix} R' \\ G' \end{pmatrix} = \begin{pmatrix} \frac{R^c}{R^c+G^c+B^c} \\ \frac{G^c}{R^c+G^c+B^c} \end{pmatrix} = \begin{pmatrix} \frac{aR^u}{aR^u+aG^u+aB^u} \\ \frac{aG^u}{aR^u+aG^u+aB^u} \end{pmatrix} = \begin{pmatrix} \frac{aR^u}{a(R^u+G^u+B^u)} \\ \frac{aG^u}{a(R^u+G^u+B^u)} \end{pmatrix} = \begin{pmatrix} \frac{R^u}{R^u+G^u+B^u} \\ \frac{G^u}{R^u+G^u+B^u} \end{pmatrix} \quad (3.16)$$

#### 3.4.2.5 Transformada do Histograma RGB

As características computadas de histogramas  $RGB$ , descrito na Seção 3.4.2.1, não possuem propriedades de invariância. Entretanto, se a distribuição dos valores dos pixels for normalizada, as características tornam-se invariantes à escala e a deslocamento em relação à variação de iluminação (Lukac and Plataniotis (2007)). Como cada canal é normalizado independentemente, as características são também invariantes à variação e a deslocamentos na cor da luz, conforme a Equação 3.17.

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} \frac{R-\mu_R}{\sigma_R} \\ \frac{G-\mu_G}{\sigma_G} \\ \frac{B-\mu_B}{\sigma_B} \end{pmatrix} \quad (3.17)$$

onde  $\mu_C$  e  $\sigma_C$  são a média e o desvio padrão da distribuição do canal  $C \in \{R, G, B\}$ , respectivamente.

### 3.4.3 Momentos

Os momentos caracterizam a distribuição de forma e intensidade para uma dada região  $\Omega$  da imagem (Hu (1962)). Uma imagem colorida, representada pela função  $I$ , está definida pelas triplas de posições  $(x, y)$ :  $I : (x, y) \mapsto$

$(R(x, y), G(x, y), B(x, y))$ . As triplas  $RGB$  representam os pontos de uma distribuição. A Equação 3.18 calcula os momentos de cor generalizados  $M_{pq}^{abc}$  definido por Gool et al. (1996).

$$M_{pq}^{abc} = \int \int_{\Omega} x^p y^q [R(x, y)]^a [G(x, y)]^b [B(x, y)]^c dx dy \quad (3.18)$$

$M_{pq}^{abc}$  é um momento de cor generalizado de ordem  $p + q$  e grau  $a + b + c$ . Assim, um número grande de momentos pode ser computado pela ordem e o grau. Entretanto, os momentos de ordem alta são menos estáveis, isto é, menos robustos às invariâncias fotométricas (Mindru et al. (2004)). Tipicamente, momentos de cor generalizados são computados até valores de ordem 1 (primeira ordem) e de grau 2 (segundo grau). Por exemplo, um momento de primeira ordem é representado por  $M_{00}^{abc}$ ,  $M_{01}^{abc}$  e  $M_{10}^{abc}$ .

Os momentos de cor  $M_{pq}^{000}$  de grau 0 não contêm informações fotométricas, enquanto momentos de ordem 0 não possuem informações espaciais. Assim, momentos de ordem 0 são invariantes à rotação, enquanto os de maior ordem não. A invariância à escala e à translação é obtida pela normalização de cada momento e pelo deslocamento dos momentos geométricos centrais, respectivamente (Jan Flusser and Zitová (2009)).

### 3.4.3.1 Momentos de Cor

Geralmente, o descritor para momentos de cor utiliza todos os momentos generalizados até o segundo grau ( $a + b + c = 2$ ) e primeira ordem ( $p + q = 1$ ). Dessa maneira, existem nove combinações possíveis de graus, para

$$abc \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (2, 0, 0), (0, 2, 0), (0, 0, 2), (1, 1, 0), (0, 1, 1), (1, 0, 1)\}$$

Combinado com as três ordens geradas ( $M_{00}^{abc}$ ,  $M_{01}^{abc}$  e  $M_{10}^{abc}$ ) o vetor possui 27 dimensões. Pela subtração da média de todos os canais, antes de computar os momentos, este descritor torna-se invariante ao deslocamento em relação à mudança de iluminação.

### 3.4.3.2 Momentos Invariantes

Com a normalização dos momentos de baixo grau e ordem é possível tornar o descritor invariante em relação a transformações fotométricas. Pela generalização de momentos de cor, os momentos invariantes envolvem apenas um único canal de cor, chamado de momento de uma banda invariante. Similarmente, há momentos invariantes para dois canais de cores, chamado de duas bandas invariantes. Para os momentos de três bandas são usados os três canais de cor. Em Mindru et al. (2004) foram computadas 3 bandas invariantes.

Assim, o vetor de características é formado por 24 momentos que são invariantes a todas as transformações fotométricas descritas na Seção 3.4.1.2.

#### 3.4.4 *Scale Invariant Feature Transform - SIFT*

O descritor SIFT é um histograma 3D que computa características das orientações do gradiente em uma região local selecionada na imagem (Lowe (2004)). A região é quantizada sobre uma grade  $4 \times 4$  e o ângulo do gradiente é quantizado em 8 orientações, resultando em um vetor de 128 dimensões para cada região. O gradiente de magnitude é calculado pela ponderação de uma janela Gaussiana sobreposta sobre a região, enfatizando os gradientes que estão mais próximos ao centro. O vetor final de características é formado pela concatenação dos histogramas de orientações de todos os *bins*. Para obter invariância à iluminação, as características são normalizadas pela raiz quadrada da soma das componentes.

Pelo fato das características serem normalizadas, as mudanças na magnitude do gradiente não têm efeito sobre o vetor final. Sob mudanças na intensidade de luz, a direção e a magnitude do gradiente são as mesmas. Além disso, as mudanças na cor da luz não causam efeito no vetor, pois a imagem de entrada é convertida para níveis de cinza. Dessa maneira, o descritor SIFT é invariante a todas as transformações fotométricas descritas na Seção 3.4.1.2. Similar às variações realizadas pela transformação dos modelos de cor, usado para computar características de histogramas (veja a Seção 3.4.2), é possível também computar SIFT usando outros modelos de cores. Bosch et al. (2007) computam características SIFT sobre os três canais do modelo *HSV*. Outra extensão similar ao SIFT é o descritor GLOH (Corso and Hager (2009)), que substitui a grade de localização cartesiana por uma grade de *log-polar*, e aplica PCA para reduzir a dimensionalidade do espaço. Ainda existem diversas modificações na literatura, entre elas, PCA-SIFT (Ke and Sukthankar (2004)), C-SIFT (Abdel-Hakim and Farag (2006)), W-SIFT (Burghouts and Geusebroek (2009)). O descritor SIFT é uma técnica de descrição que emprega um detector de região para computar características locais. O detector aproxima o filtro Laplaciano-da-Gaussiana (LoG) a uma diferença-de-Gaussianas (DoG). Este grupo de descritores, juntamente com o modelo *bag-of-visual-features*, são discutidos na próxima seção.

### 3.5 Características Locais e o Modelo Bag-of-Visual-Features

Inspirados em estudos sobre pré-atenção distintiva de Julesz (1962) e Neisser (1964), Schmid and Mohr (1997) sugeriram que apenas a informação local poderia ser suficiente para descrever o conteúdo de imagens. Schmid and Mohr (1997) compararam diferentes detectores de região para descrever a repetibilidade dos pontos de interesse usados na representação de imagens. A informação de conteúdo local foi medida considerando-se a entropia da distribuição de valores em níveis de cinza computados em torno dos pontos detectados. Tais pontos geralmente são localizados em cantos e fazem a categorização de objetos robusta a diversas transformações.

Em contrapartida, vários trabalhos foram propostos para reconhecer objetos por meio de invariância geométrica (Taubin and Cooper (1992); Rothwell (1994)) e características globais (Niblack et al. (1993); Pontil and Verri (1998); Schiele and Crowley (2000)). Entretanto, a variação em condições de iluminação, a sobreposição de objetos e mudanças na posição da câmera têm um impacto significativo sobre a aparência das imagens. Um paradigma de representação que tem recebido muita atenção para lidar com estes problemas é a descrição de imagens usando características locais (Jia et al. (2008)). Tipicamente, esta descrição cria uma representação em forma de histogramas sobre um vocabulário discreto de características locais, isto é, os vetores computados dos pontos salientes são atribuídos aos *bins* da distribuição de acordo com a partição do espaço de características.

A ideia em representar as características em uma distribuição foi motivada pela analogia aos modelos *bag-of-words*, usados em *mineração de texto*. Este modelo obteve um grande sucesso em sistemas de recuperação de documentos, e mais tarde, Zhu et al. (2002) adaptaram esta abordagem para categorização visual. Os autores estudaram a quantização de vetor de pequenas janelas de imagens, a princípio quadradas, chamadas de *keyblocks*. Daí a diante, devido à falta de uma nomenclatura padrão na literatura, novas abordagens usando quantização de vetores de características locais foram propostas como *bag-of-visual-features* (Sivic and Zisserman (2003)), *bag-of-visual-words* (Corso and Hager (2009)), *bag-of-keypoints* (Ommer and Buhmann (2010)), *bag-of-features* (Lazebnik et al. (2006)) e geração de *codebooks* visuais (Jurie and Triggs (2005); Burghouts and Geusebroek (2009); Lazebnik and Raginsky (2009)). No presente trabalho, a terminologia adotada é *bag-of-visual-features* - (BoVF).

Quantificar imagens estatisticamente como histogramas sobre um vocabulário discreto de características locais foi aplicado inicialmente em tarefas

de classificação de textura (Schmid and Mohr (1997); Leung and Malik (2001); Cula and Dana (2001); Ojala et al. (2002); Hayman et al. (2004); Lazebnik et al. (2005, 2006); Varma and Garg (2007); Lazebnik and Raginsky (2009); Varma and Zisserman (2009)). A classificação de texturas possui diferentes escopos comparada à categorização geral. Dessa maneira, é comum encontrar diferentes propostas na literatura. Em (Leung and Malik (2001); Cula and Dana (2001)) respostas de descritores locais (filtros) são clusterizadas e o reconhecimento de textura é feito pela medida mais próxima calculada pelo teste do  $\chi^2$ . Fergus et al. (2003) propuseram um método que explora um modelo probabilístico que combina probabilidades dos *patches* de imagens em relação à escala e posição. Nas abordagens propostas por (Ojala et al. (2002); Hayman et al. (2004); Varma and Garg (2007); Varma and Zisserman (2009)), cada imagem é representada como um histograma sobre um dicionário comum, seguida pela comparação de histogramas de imagens. Outra abordagem encontrada em Lazebnik et al. (2005) usa um dicionário para cada imagem e a representa como uma assinatura visual. Este dicionário é formado pelo agrupamento dos vetores de características e definido pelo número de *bins* na distribuição. Assim, cada imagem é descrita e associada ao centróide mais próximo de acordo com o particionamento do espaço de características.

O diagrama da figura 3.1 apresenta os passos para construir o modelo BoVF. O processo inicia pela seleção de pontos de interesse na imagem que são usados para computar os vetores de características em um certo raio. Existem diversas estratégias e algoritmos propostos na literatura (Mikolajczyk et al. (2005)), podendo ser divididas em dois grupos: amostragem densa e detectores de região (detecção esparsa). A amostragem densa é uma maneira simples de obter regiões locais uniformemente amostradas em uma imagem. As regiões são determinadas por uma grade regular com passo correspondente ao tamanho do raio da região. Dessa maneira, se o raio é pequeno, a grade é amostrada com uma quantidade grande de pontos, caso contrário, os pontos são mais espaçados e resulta em um número menor de amostras. A amostragem densa é computacionalmente cara pelo número grande de pontos e, conseqüentemente, o número de vetores de características computados é alto. Esta estratégia é utilizada para casos em que as imagens são pequenas e apresentam baixo contraste.

Diferentemente da amostragem densa, existem algoritmos para encontrar as regiões locais. Os detectores de pontos de interesse são usados para extrair estruturas mais salientes nas imagens, muitas vezes utilizando critérios que tornam a descrição invariante à mudança de escala e à transformação afim. Harris and Stephens (1988) propuseram o detector *Harris* baseado na matriz de auto-correlação, usada para detecção de pontos e descrição da região. Este

detector é invariante apenas a rotação. Mikolajczyk et al. (2005) propuseram o detector de regiões *Harris-Laplace*, invariante à rotação e mudanças de escalas. Os pontos são detectados pela função *Harris* de escala adaptada e selecionada no espaço de escala por meio do operador *Laplaciano-da-Gaussiana* (LoG). Esta característica de seleção é determinada pela proposta de Lindeberg. Lowe (2004) aproxima o *Laplaciano* com diferença-de-Gaussianas (DoG). Matas et al. (2004) propuseram o detector de regiões *maximally stable extrema* extraído com o algoritmo de segmentação *watershed*. Uma revisão do estado-da-arte em detectores pode ser encontrada em Mikolajczyk et al. (2005).

As regiões circulares obtidas pelos detectores são usadas para computar os vetores de características. Uma lista de descritores é apresentada na Seção 3.2. Depois da seleção e descrição, o espaço de características dos pontos é quantizado contra um vocabulário visual por meio do particionamento do espaço de características. Uma estratégia de particionamento que passou ser predominante e permite a representação em forma de histograma é o algoritmo *k-means* (MacQueen (1967)). Não somente este esquema de particionamento preenche o espaço de características suficientemente para permitir o agrupamento dos vetores, mas ele também controla o número de caixas na representação por meio da definição do número de centróides com agrupamento não supervisionado de vetores de características locais (Jurie and Triggs (2005); Burghouts and Geusebroek (2009); Lazebnik and Raginsky (2009)). Assim, um descritor é atribuído ao elemento *codebook* (*codevector*) que está mais próximo no espaço Euclidiano. O resultado é um histograma normalizado até 1 que representa o modelo BoVF.

### 3.6 Considerações Finais

O problema central em muitas tarefas de análise de imagens e visão computacional está relacionada em considerar a descrição efetiva de imagens. Neste capítulo, foi apresentada uma revisão sobre as técnicas de descrição. Determinar qual representação será aplicada sobre as imagens é ainda uma tarefa difícil devido a necessidade de vários testes para se descobrir o quão discriminante tais representações são. Na Seção 3.2 são introduzidas duas abordagens para representação de imagens, dentre elas, *contorno* e *região*. Para esta dissertação, é explorado apenas a representação por região, pois os conjuntos contêm imagens em que a informação de *textura* e *cor* são mais salientes. Entretanto, é possível empregar outras técnicas para descrever imagens, tais como as baseadas em forma dos objetos (Chi and Leung (2008); Datta et al. (2008); Costa and Cesar Jr. (2009)). Na Seção 3.3 foram descritas duas técnicas tradicionais no contexto de textura, matriz de co-ocorrência e filtros de



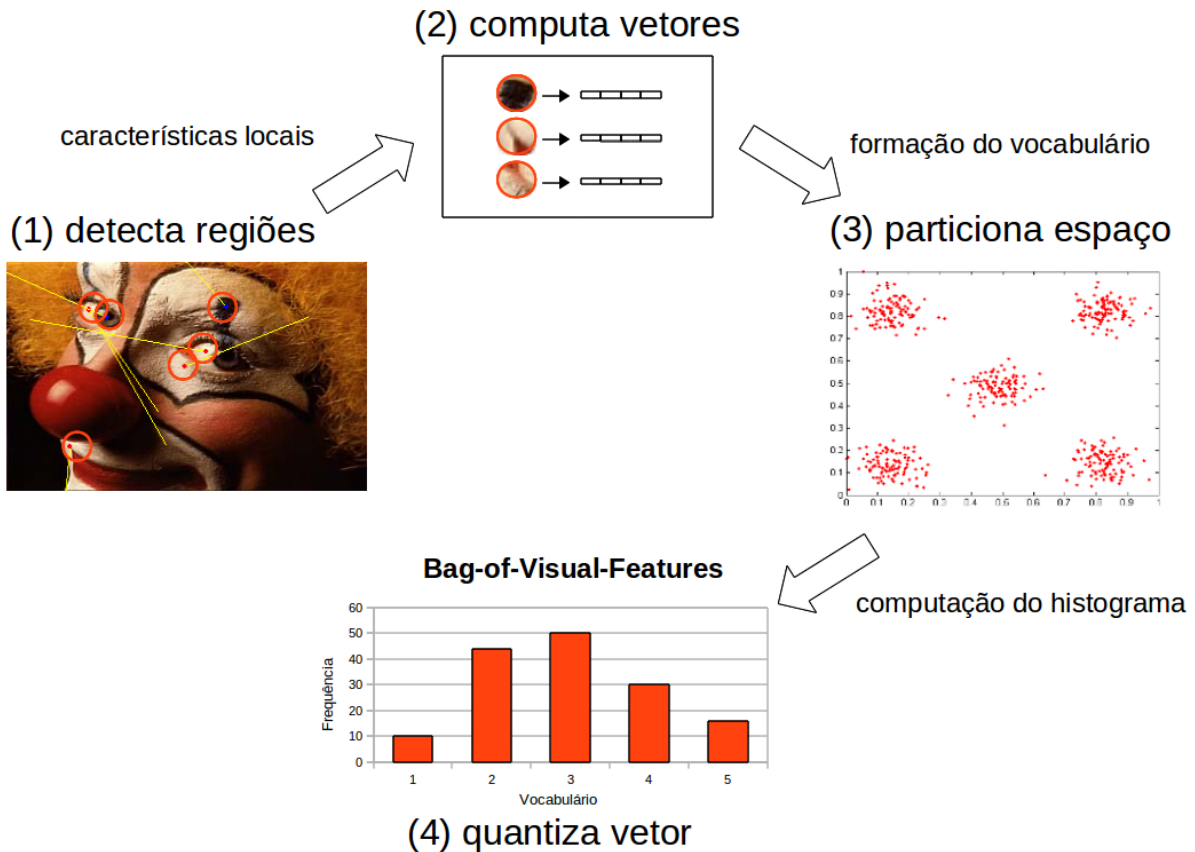


Figura 3.1: Fluxograma para construção do modelo *bag-of-visual-features* - (BoVF).

Gabor.

Para a proposta de descrição de imagens baseada em cor, foram apresentados, na Seção 3.4, três grupos de técnicas: *histogramas*, *momentos de cor* e descritores baseado em *Scale Invariant Feature Transform* - SIFT. No entanto, nos últimos anos a descrição de imagens por características puramente computada de informações locais tem recebido uma grande atenção (Seção 3.5). Esta abordagem representa imagens estatisticamente como histogramas sobre um vocabulário discreto de características locais. Este vocabulário é definido pelo particionamento do espaço de características, tipicamente baseado em agrupamento não supervisionado. Esta abordagem foi inspirada nos modelos *bag-of-words* (BoW) e adaptada para categorização visual, chamado de *modelo bag-of-visual-features* (BoVF). O alto desempenho atribuído a este modelo está ligado à tolerância à variações de iluminação, distorções, mudanças de escala, oclusão de objetos e mudanças na posição da câmera. Isso faz com que tais características sejam usadas em muitas aplicações de visão computacional, tais como recuperação de imagens, reconhecimento de objetos e texturas, *matching* de imagens e localização por pontos (*landmarks*).



---

# UMA NOVA ABORDAGEM PARA EXPLORAÇÃO VISUAL DO ESPAÇO DE CARACTERÍSTICAS APLICADA À ANÁLISE DE IMAGENS

---

Este capítulo apresenta a abordagem proposta para exploração visual do espaço de características. A Seção 4.1 descreve a metodologia aplicada à análise de imagens. Na Seção 4.2, experimentos são executados em seis conjuntos de imagens. Nessa mesma seção são descritos os conjuntos de imagens, e discutidos os detalhes da configuração de descritores e classificadores. Este capítulo também discute os resultados de cada experimento. Finalmente, as considerações finais do capítulo são feitas na Seção 4.3.

## 4.1 *Metodologia Proposta*

O objetivo deste projeto de mestrado é criar uma abordagem para analisar visualmente diferentes espaços de características. Cada espaço é apresentado em uma representação gráfica de baixa dimensão usando técnicas de visualização baseadas em projeção. Uma técnica de projeção mapeia dados de alta dimensionalidade em espaços visuais bidimensionais, buscando preservar as relações de distâncias definidas no espaço original. Basicamente, os vetores de características computados são visualizados em uma representação que revela as relações de distância entre as imagens analisadas. Esta repre-

sentação visual é usada para determinar se tais características representam apropriadamente o conjunto de imagens, de acordo com o ponto de vista do especialista. Se as relações de similaridade correspondem ao que é esperado pelo especialista, isto é, imagens similares são posicionadas proximamente e imagens dissimilares são posicionadas distantemente, o conjunto de características representa apropriadamente o conjunto de dados. Este conjunto, portanto, pode beneficiar sistemas de análise de imagens baseados em características. Caso contrário, os parâmetros podem ser alterados ou outra técnica de extração pode ser empregada, produzindo um novo conjunto de características que pode ser novamente explorado, de tal forma que a representatividade deste novo conjunto possa ser avaliada. Isto fornece uma abordagem visual que ajuda usuários a construir modelos mais robustos aplicados à análise de imagens.

O diagrama da Figura 4.1 resume a abordagem proposta neste trabalho. Inicialmente, um conjunto de características de imagens é computado (1) e, em seguida, projetado em um espaço 2D (2). Para avaliar a qualidade de tais projeções, o especialista analisa visualmente a projeção, com o apoio do *índice da silhueta* (3). Assim, de acordo com a qualidade da representação visual, o conjunto de características pode, então, ser usado em outras tarefas, tais como classificação, recuperação ou mineração de dados. Caso uma quantidade significativa de pontos (instâncias) esteja visualmente mal posicionada, os parâmetros ou as técnicas de extração podem ser alterados (4), e o processo repetido. Este processo é executado até convergir para um conjunto de características que reflita em uma boa separabilidade das classes na projeção.

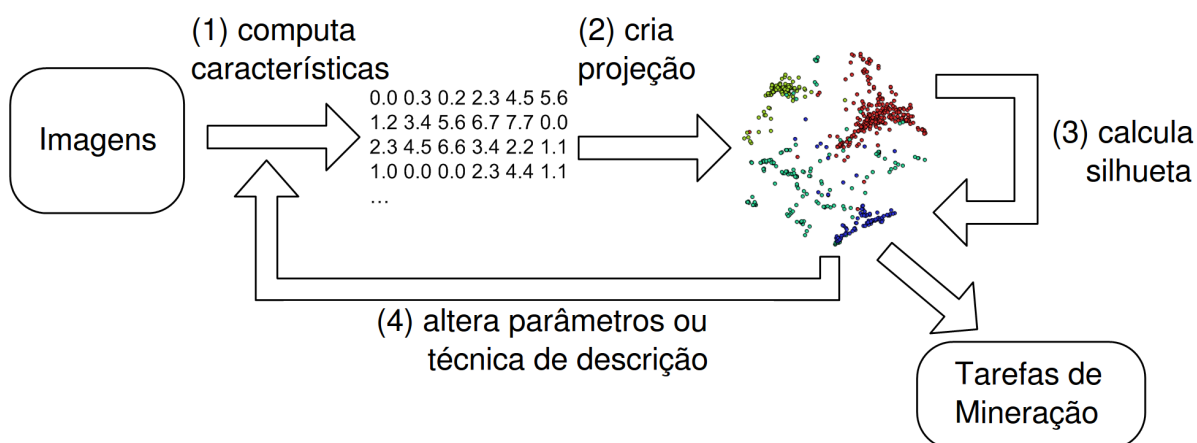


Figura 4.1: Proposta de exploração visual do espaço de características.

A visualização do espaço de características pode fornecer informações importantes sobre a estrutura dos dados no espaço original. Assim, a exploração visual por meio de projeções apóia especialistas a realizar tarefas para:

- Descoberta de relacionamentos *inter-classe*, tais como identificação de

classes similares e dissimilares.

- Avaliação da distribuição e discriminação de propriedades de classes individuais.
- Identificação de classes com dispersão de pontos, possivelmente prejudicando a discriminação de outras classes.
- Avaliação de um ponto de vista geral sobre a qualidade na formação de grupos no espaço de características (separação e coesão).

A análise de tais questões pode permitir que especialistas predizam a qualidade da classificação, usando as representações para inspecionar relacionamentos de similaridade e a qualidade visual do espaço de características. Descobrir padrões em relações *inter-classe* pode ser útil para sistemas de análise de imagens de forma a convergir mais rapidamente para conjuntos de características mais efetivos e também na construção de classificadores mais robustos. Para este fim, é fundamental a escolha de técnicas de projeção que preservam o mais fielmente possível as relações de similaridade.

Formalmente, dado um conjunto  $X = \{x_1, x_2, \dots, x_n\}$  de  $n$  imagens representadas em um espaço de características  $m$ -dimensional, e  $\delta$  uma medida de dissimilaridade entre dois vetores de características  $m$ -dimensional, e  $d$  uma distância (normalmente Euclidiana) entre dois elementos do espaço projetado. Uma projeção multidimensional pode ser interpretada como uma função  $f : \mathbb{R}^m \rightarrow \mathbb{R}^p$  (com  $p = \{1, 2, 3\}$ ) em que se busca fazer  $|\delta(x_i, x_j) - d(f(x_i), f(x_j))|$  como o mais próximo de zero possível,  $\forall x_i, x_j \in X$  (Tejada et al. (2003)).

Atualmente, existe um grande número de técnicas de projeção considerando diferentes aspectos da distribuição dos dados (Paulovich (2008)). Como discutido na Seção 2.5, a redução de dimensionalidade de espaços em alta dimensão para representações 2D ou 3D causa uma perda inevitável de informação. Dessa maneira, o interesse recai sobre técnicas que preservam as relações espaciais entre as classes, buscando preservar ao máximo na projeção as relações de similaridade presente nos dados originais. Para cumprir esta tarefa, a técnica de projeção *Classical Scaling*, uma variante da MDS, é considerada a melhor escolha (Cox and Cox (2001)).

A técnica *Classical Scaling* é definida por uma matriz de centragem dupla entre todos os pares de instâncias, calculando-se a decomposição espectral para recobrir as coordenadas cartesianas dos elementos no espaço visual. Formalmente, é possível provar que se a função de distância é Euclidiana, o espaço projetado apresenta o menor desvio quadrático médio do espaço original entre todos os possíveis espaços reduzidos. Isto é, entre todas as matrizes  $\mathbf{A}_{n \times p}$ , a projeção  $\mathbf{Y}_{n \times p}$  é aquela que minimiza  $\|\mathbf{X} - \mathbf{A}\|^2 = \sum_{i,j} (x_{i,j} - a_{i,j})^2$  (Mardia et al. (1995)). Assim, a *Classical Scaling* preserva o máximo possível as

relações de distância entre as instâncias de dados quando eles são projetados no espaço visual (Cox and Cox (2001)).

A metodologia aqui proposta foi implementada usando a ferramenta PEx-Image. Além de adaptar as funcionalidades existentes na PEx-Image foi adicionado à ferramenta o índice da silhueta para avaliar os espaços de características. Além disso, foi criado um novo módulo para descrever imagens. O desenvolvimento de uma nova versão da ferramenta foi implementada e disponibilizada no repositório do grupo VICG/LCAD<sup>1</sup>. A avaliação da abordagem proposta será apresentada na próxima seção.

## 4.2 Avaliação Experimental

Com o objetivo de avaliar a abordagem proposta nesta dissertação, experimentos são executados em seis conjuntos de imagens. Os conjuntos de imagens empregados são descritos na Seção 4.2.1. Na Seção 4.2.2 são discutidos detalhes da configuração de descritores e classificadores. Finalmente, os resultados são apresentados na Seção 4.2.3.

### 4.2.1 Conjuntos de Imagens

Com o intuito de realizar uma avaliação mais abrangente, três categorias de bases de imagens são consideradas: *texturas*, *imagens médicas* e *cenais naturais (outdoor)*. Conjuntos com presença de texturas são: **Brodatz** (Brodatz (1966)), **Textured Surfaces** (Lazebnik et al. (2005)) e **KTH-TIPS** (Hayman et al. (2004)). Esta abordagem é também testada em uma base de imagens médicas, chamada **ImageCLEF2006** (Clough et al. (2007)). Além disso, cenais naturais são testadas com os conjuntos **Corel-1000** (Li and Wang (2003)) e **Caltech** (Fergus et al. (2003)). Estes conjuntos são descritos nas próximas subseções.

#### 4.2.1.1 Imagens com Textura

**O álbum de Brodatz** (Brodatz (1966)) é um dos conjuntos de imagens mais conhecidos para avaliar algoritmos de reconhecimento de texturas. Embora ele contenha 112 classes, poucas publicações descrevem resultados sobre o banco de imagens inteiro. Cada classe é representada por uma imagem dividida em nove sub-imagens, mais a textura original, totalizando 10 imagens. Exemplos de três classes são mostradas na Figura 4.2. A variação de escalas e padrões geométricos das texturas de Brodatz, faz deste conjunto uma boa plataforma para testar descritores baseados em textura.

---

<sup>1</sup>Acessado apenas por membros em <http://svn.lcad2.icmc.usp.br/>

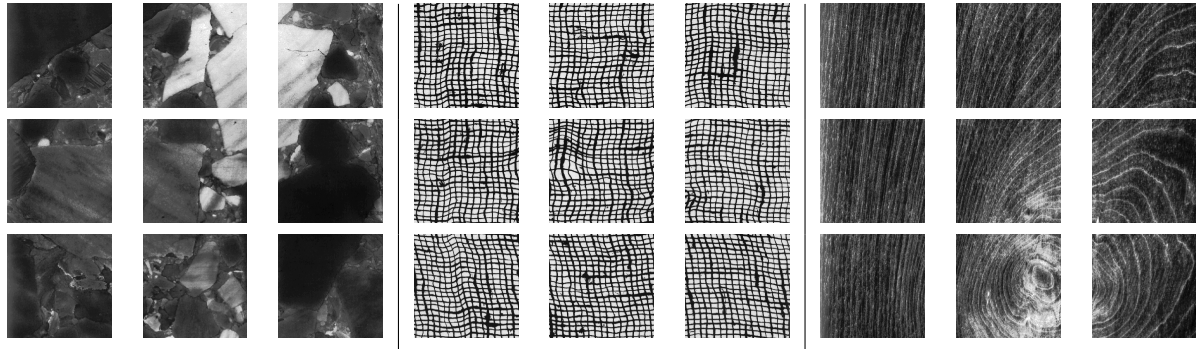


Figura 4.2: Exemplo de três imagens de texturas de Brodatz. Cada imagem é dividida em nove sub-imagens sem sobreposição.

**O conjunto Textured Surfaces** (Lazebnik et al. (2005)) contém 25 classes de texturas com 40 imagens por classe, totalizando 1000 imagens. Este conjunto busca superar críticas direcionadas às texturas de Brodatz, relacionadas com a baixa variação intra-classe, a uniformidade dos padrões de textura e a planaridade das superfícies de textura. Assim como Brodatz, as texturas são disponibilizadas em diferentes escalas. Entretanto, este conjunto inclui deformações não rígidas, variação em condições de iluminação e mudanças no ponto de vista da câmera. A Figura 4.3 apresenta quatro imagens por classe, cada uma delas mostrando um superfície de textura vista de diferentes orientações.

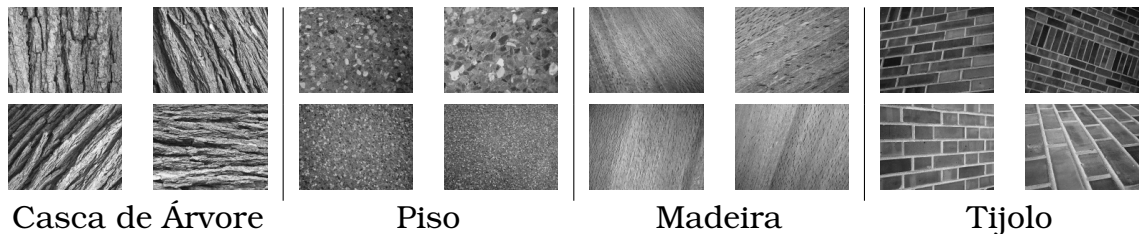


Figura 4.3: Quatro exemplos de texturas do conjunto *Textured Surfaces*. Este conjunto supera críticas sobre *Brodatz* devido à variação em condições de iluminação, mudanças no ponto de vista da câmera, além da não planaridade nas superfícies de textura, como a classe *casca de árvore*.

**O conjunto KTH-TIPS** (Hayman et al. (2004)) consiste de 10 classes de textura colorida. As imagens são produzidas em 9 diferentes escalas, variando de 0.5 a 2. Cada escala é visualizada sob três direções de iluminação e três poses, resultando um total de 9 imagens por escala, e 81 imagens por material. Algumas imagens com variações na escala e iluminação são mostradas na Figura 4.4. Este conjunto torna-se interessante para estudos que avaliam a contribuição da cor presente nas texturas.

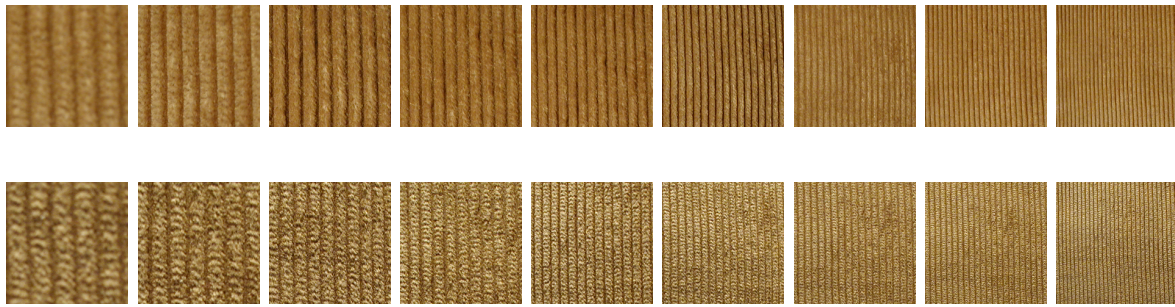


Figura 4.4: Exemplos da classe *tecido* do conjunto de materiais *KTH-TIPS*. As mudanças de escala e iluminação aumentam a variação intra-classe. As duas linhas de imagens mostram a mudança de iluminação para o mesmo material. As colunas representam as nove escalas do conjunto.

#### 4.2.1.2 *Imagens Médicas*

**O conjunto ImageCLEF2006** (Clough et al. (2007)) consiste de imagens de raio-X usadas para avaliar sistemas de recuperação de imagens baseada em conteúdo (CBIR). Este conjunto surgiu em 2003 como uma competição para avaliar tarefas de mineração de texto. Em 2004, foi adicionada ao conjunto uma trilha para avaliar sistemas de recuperação de imagens médicas. Não somente a recuperação pode ser feita por meio da análise de imagens, mas também de acordo com suas anotações; ou pela combinação de características de imagens e anotações. Exemplos de imagens são mostradas na Figura 4.5. Devido ao desbalanceamento no número de instâncias por classe, para este trabalho são usadas apenas categorias em que o número de amostras varia entre 50 e 200, totalizando 3.870 imagens de raio-X divididas em 35 classes.

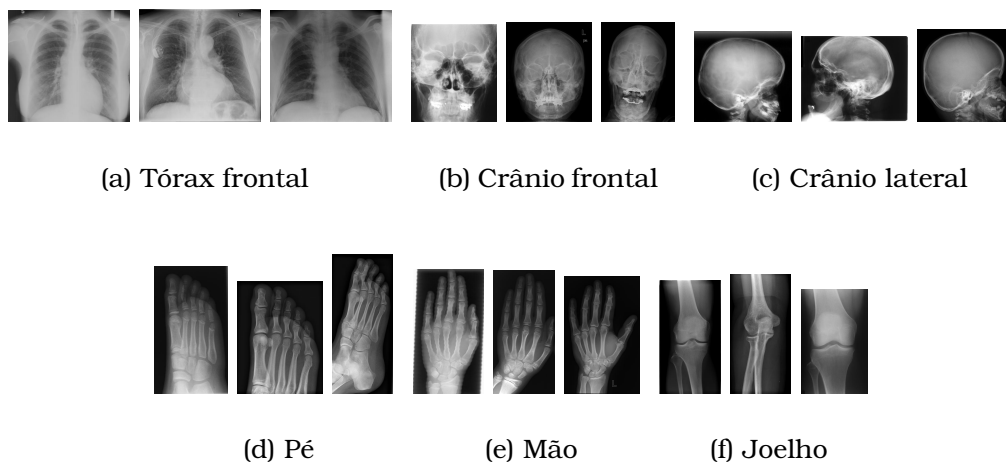


Figura 4.5: Exemplos de imagens do conjunto *ImageCLEF2006*. Este conjunto é caracterizado pela alta variação intra-classe.



### 4.2.1.3 *Imagens de Cenas Naturais*

O conjunto **Corel-1000** (Li and Wang (2003)) contém 1000 imagens de cenas naturais e objetos artificiais separados em 10 classes. Ele é usado amplamente pela comunidade de recuperação de imagens. A proposta deste conjunto é avaliar o desempenho de sistemas de recuperação de imagens e obter o entendimento de como as técnicas de descrição funcionam. Uma imagem para cada classe é apresentada na Figura 4.6. Embora esta base possua alta variação intra-classe, o fundo possui predominância em termos de cor. Dessa maneira, estratégias globais para descrição de imagens conseguem representar apropriadamente o conjunto.

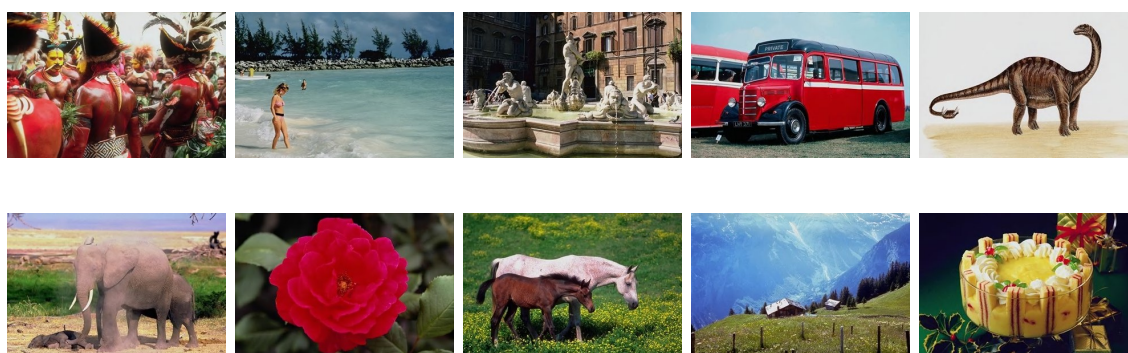


Figura 4.6: Exemplos de imagens do conjunto *Corel* separadas em 10 classes. Apesar da presença de variação intra-classe, este conjunto possui alta correlação em termos de fundo.

O conjunto **Caltech** (Fergus et al. (2003)) contém seis classes e um total de 3812 imagens. As classes são: avião (1074 imagens), edificação (750 imagens), carro (526 imagens), face (450 imagens), folha (186 imagens) e motocicleta (826 imagens). Este conjunto é considerado desafiador pois ele inclui imagens com alta variação intra-classe e oclusão de fundo. Além disso, há uma grande variação na informação do fundo das imagens. O objetivo em abordar cenas naturais nos experimentos é avaliar as técnicas de cor e aplicar o modelo *bag-of-visual-features*. Algumas imagens deste conjunto são apresentadas na Figura 4.7.

## 4.2.2 *Configuração dos Experimentos*

Nesta seção são dados detalhes sobre os detectores de pontos de interesse usados como estágio inicial para computar características locais e construir o modelo *bag-of-visual-features*. Ainda nesta seção são apresentados detalhes da parametrização de descritores e classificadores.

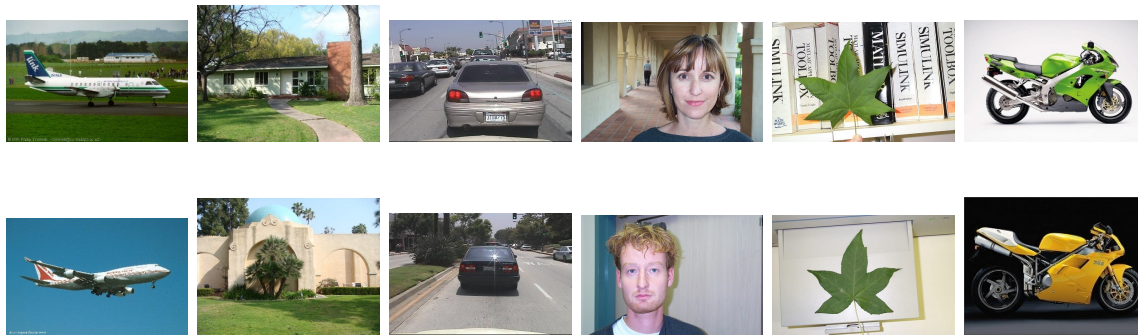


Figura 4.7: Dois exemplos para cada classe de imagens do conjunto *Caltech*. Ao contrário do conjunto do *Corel*, este banco possui classes em que a informação de fundo não é correlata.

#### 4.2.2.1 Descrição de Imagens

Nos experimentos a seguir, são empregados os métodos tradicionais de descrição baseados em textura, juntamente com descritores de cor. Para imagens que apresentam informação de cor, são também computadas características locais quantizadas sobre um vocabulário visual para construir o modelo *bag-of-visual-features* - (BoVF) (descrito na Seção 3.5). Os pontos de interesse são selecionados com o detector de região Harris-Laplace (Mikolajczyk et al. (2005)) e amostragem densa (Jurie and Triggs (2005)), dependendo da natureza das imagens. Conforme texturas tendem a ser mais regulares e cenas naturais apresentem diferentes objetos de interesse e fundos, é comum selecionar pontos com amostragem densa para texturas e empregar detectores de região para cenas naturais. Existem diversos outros detectores de região que podem ser utilizados, tais como diferença-de-Gaussianas (DoG) (Lowe (2004)), *maximally stable extrema* (Matas et al. (2004)) (veja a Seção 3.5). O uso do detector Harris-Laplace é motivado pela estabilidade sob invariâncias à transformações geométricas e mudanças de iluminação, resultando em características mais discriminatórias.

Para conjuntos de imagens em que a presença de textura é mais evidente - **Brodatz**, **Textured Surfaces**, **ImageCLEF2006** e **KTH-TIPS** - propriedades físicas são capturadas por meio da análise de textura, e para conjuntos que contêm a informação de cor - **KTH-TIPS**, **Corel-1000** e **Caltech** - são exploradas a contribuição das características globais e características locais. Características locais de cor são extraídas sobre uma região de 21 pixels ao redor dos pontos de interesse. De acordo com Mikolajczyk et al. (2005), esta área corresponde ao máximo de escala do Lapaciano-da-Gaussiana. No entanto, cada imagem é representada por uma quantidade diferente de pontos de interesse (*keypoints*). O modelo BoVF é empregado para quantizar o espaço de características e obter um vetor de características com tamanho fixo.

Depois de detectar tais pontos e computar os vetores, o vocabulário visual é criado pelo algoritmo de agrupamento *k-means* (MacQueen (1967)). O algoritmo *k-means* é considerado uma alternativa simples para particionar o espaço de características, pois ele não requer um passo de treinamento para aprender o vocabulário visual. Embora o particionamento do espaço de características seja não supervisionado, o algoritmo *k-means* pode convergir apenas para um ótimo local. Apesar desta limitação, ele preenche o espaço de características de forma satisfatória para permitir o agrupamento dos vetores, além de controlar o número de *bins* na representação por meio da definição do número de centróides. Aqui, o parâmetro  $k$  do vocabulário varia entre 50 e 300 e o número de *keypoints*, varia de 50.000 a 150.000. Neste trabalho, análise de componentes principais (PCA) (Jolliffe (2002)) é também usada para reduzir a dimensionalidade como etapa intermediária.

Para os resultados apresentados neste capítulo são usados os descritores histogramas RGB, momentos invariantes de 3 bandas (Mindru et al. (2004)) e SIFT (Lowe (2004)) (descritos na Seção 3.4). Para computar o vetor SIFT é usada a versão disponibilizada por Lowe (Lowe (2004)). Características de texturas são computadas pelas técnicas matriz de co-ocorrência e filtros de Gabor. No primeiro, são extraídas características com 5 medidas estatísticas (energia, entropia, inércia, momento de diferença inversa e correlação), 5 distâncias (de 1 até 5) e 4 orientações ( $0^0$ ,  $45^0$ ,  $90^0$  e  $135^0$ ), totalizando um vetor com 100 dimensões. No segundo, são computadas 16 características usando energia da resposta dos filtros de Gabor criados com 4 orientações ( $0^0$ ,  $45^0$ ,  $90^0$  e  $135^0$ ) e 4 escalas. Além disso, experimentos são também realizados sobre texturas coloridas usando características de Gabor extraídas dos canais  $R$ ,  $G$ ,  $B$ , produzindo um vetor de 48 dimensões. Para o conjunto **ImageCLEF2006**, o espaço é explorado com características computadas de transformadas wavelets de Haar em 4 níveis de decomposição, produzindo 1024 dimensões no vetor.

#### 4.2.2.2 Classificação

Existem diversos esquemas de classificação propostos para categorização de dados. Embora a superioridade de *Support Vector Machines (SVMs)* (Vapnik (1995)) para classificação de dados foi claramente demonstrada na literatura, este trabalho não pretende alcançar o melhor desempenho de classificação. Ao invés disso, busca confirmar a seguinte hipótese: conforme a qualidade das representações visuais melhora, o desempenho de classificação também aumenta. Dessa maneira, *k-means* (MacQueen (1967)) e *Nearest Neighbors (K-NN)* (Cover and Hart (1967)) com distância Euclidiana são usados para classificação de imagens. A implementação *LibSVM* (Chang and Lin (2001))

é usada para treinar o classificador SVM. Para o classificador K-NN, são usados quatro valores de  $K = \{10, 30, 50, 70\}$ , dependendo do número de instâncias do conjunto de imagens. Para o *k-means*, o número de grupos é definido como sendo igual ao número de classes de cada conjunto. Aqui, SVM é empregado com dois *kernels* para categorização multiclasse: (1) linear e (2) função de base radial (RBF). Para cada tamanho de vocabulário, o parâmetro custo é otimizado usando validação cruzada 10 dobras com parâmetro de escala variando entre  $2^{-10}$  e  $2^{10}$ , tomando 10 como passo em escala logarítmica.

### 4.2.3 Resultados

#### **Experimento 1**

Um conjunto de 70 imagens do álbum de **Brodatz**, agrupada em 7 classes com 10 imagens cada, é empregado no experimento. Características de texturas são extraídas usando filtros de Gabor, com 4 escalas e 4 orientações. A visualização do espaço de características é mostrada na Figura 4.8. As imagens são representadas como círculos e coloridas de acordo com a classe a que pertencem. Vale ressaltar que a informação disponível da classe não é usada para gerar a projeção. Neste caso, uma imagem de cada classe é colocada como um ícone próximo ao grupo correspondente. A projeção revela cinco classes bem separadas, enquanto as outras duas estão misturadas (canto superior à direita). Embora possua uma forte relação inter-classe entre estas duas classes, isto indica que outro conjunto de características possa ser avaliado. Assim, cabe ao usuário decidir se tais classes não deveriam ser consideradas uma única classe. Se for este o caso, o conjunto de características pode ser considerado um bom conjunto, caso contrário, um novo conjunto deve ser avaliado.

O experimento é então estendido com a adição de 3 classes de textura (nomeadamente mármore e duas classes de arame), e a projeção resultante é mostrada na Figura 4.9(a). Para facilitar o entendimento, imagens de tais classes são exibidas como ícones ao lado de seu grupo. O mesmo conjunto de parâmetros é usado para extrair o conjunto de características. A classe mármore, cuja amostras são destacadas em vermelho, aparece espalhada por toda a projeção. Um usuário sem muito conhecimento pode facilmente concluir que o conjunto de características escolhido não é adequado para representar a classe mármore, o que provavelmente corresponde ao resultado esperado por um usuário experiente. Imagens da classe mármore, normalmente, apresentam elementos de textura não uniforme, que variam em tamanho e orientação. Essa característica é ilustrada na Figura 4.9(b), que mostra a exploração por meio da ampliação (*zoom-in*) da área referente à classe mármore correspon-

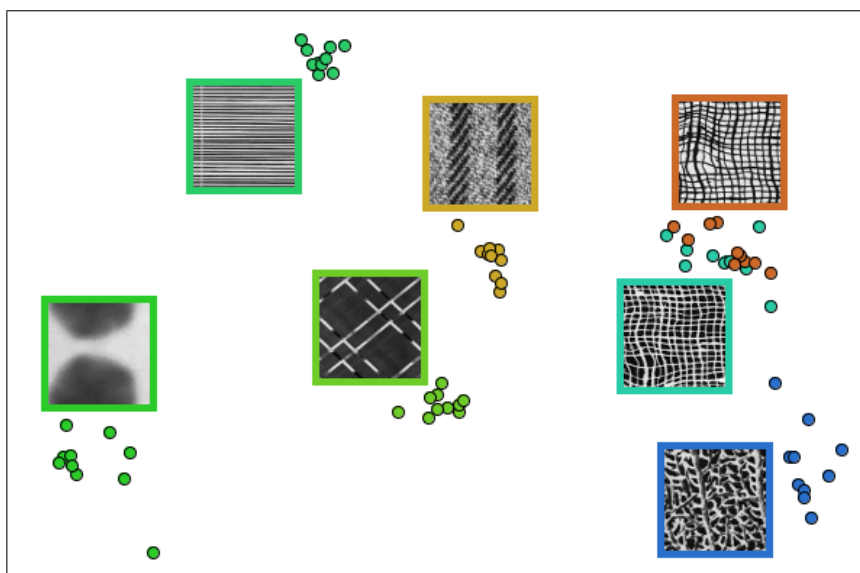


Figura 4.8: Análise de características de Gabor com 7 classes de texturas do álbum de *Brodatz*. Cinco classes de imagens são bem separadas, mas 2 são misturadas entre si. Isto indica que características extraídas não representam efetivamente este conjunto de imagens.

dente na Figura 4.9(a). Embora sejam variados os parâmetros para diferentes orientações e escalas, é improvável que todos os padrões desta textura sejam capturados. Isso está relacionado com a presença de variação intra-classe.

Para as duas classes arames, as amostras projetadas estão misturadas, como pode ser observado no canto superior esquerdo da Figura 4.9(a). Para alcançar melhor separabilidade, apenas características com orientação  $90^\circ$  são empregadas (o usuário ignorou as outras três orientações, definindo suas ponderações para zero) e uma nova projeção é computada. A visualização resultante é mostrada na Figura 4.10, com a separação das duas classes que antes estavam misturadas. Isto é esperado, uma vez que os padrões de textura para orientação  $90^\circ$  são diferentes para cada classe. No entanto, quando comparada com a projeção gerada anteriormente, considerando-se características diferentes, as demais classes estão mais espalhadas. Neste caso, o usuário pode concluir que a orientação para  $90^\circ$  é eficaz na separação destas duas novas classes de imagens, mas a coesão das outras classes é prejudicada. Este resultado não é fácil de se obter, sem visualizar o espaço de características. Em termos do índice da silhueta, as características empregadas para gerar a projeção da Figura 4.10 produz melhores valores quando comparadas às características empregadas para projetar o espaço mostrado na Figura 4.9(a). No entanto, a diferença é pequena, 0,429 para o primeiro e 0,474 para o segundo, que coincide com a inspeção visual.

Quaisquer que sejam as orientações escolhidas no experimento anterior, é possível perceber que as características de Gabor não conseguem separar as amostras das duas classes de linha do conjunto de Brodatz. Suas amos-

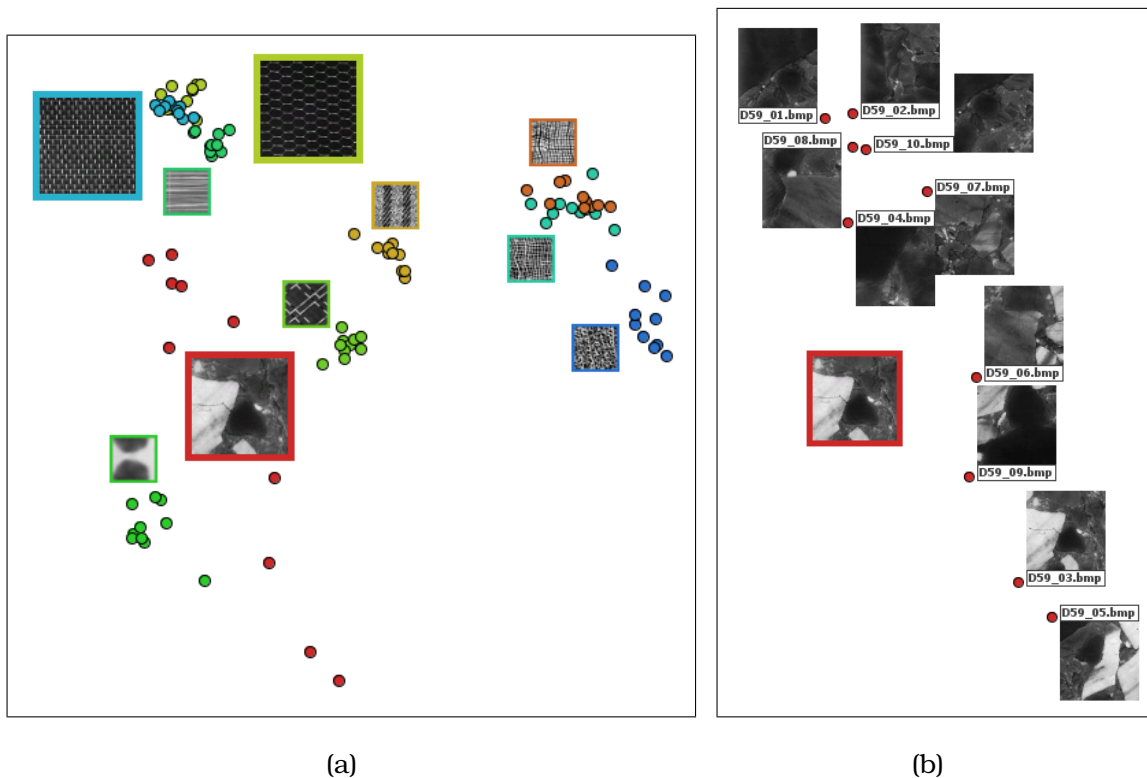


Figura 4.9: Projeção de 10 classes de texturas de *Brodatz* usando características de Gabor. Devido a baixa uniformidade no padrão da classe mármore, é facilmente verificado, mesmo por um usuário não-especialista ou sem conhecimento *a priori*, que o conjunto de características falha ao representar as imagens. Isso fornece ao especialista novas hipóteses de exploração do espaço.

tras são coloridas em verde e marrom e estão posicionadas no canto superior direito da Figura 4.8. Elas também estão presentes nas Figuras 4.9(a) e 4.10.

Na verdade, essas duas classes apresentam padrões de textura muito semelhantes e uma inspeção mais minuciosa revela apenas uma ligeira variação na intensidade dos pixels. Neste caso, as características extraídas com filtros de Gabor não foram eficazes para separar tais classes. Em busca de uma melhor separabilidade, é investigado um novo conjunto de características extraído com a técnica matriz de co-ocorrência, como descrito na Seção 3.3.1. A projeção resultante é mostrada na Figura 4.11. As duas imagens ampliadas representam as duas classes distintas de linho, mostradas à direita. Claramente, as amostras na cor verde e marrom são agora bem separadas. A inspeção visual indica que a matriz de co-ocorrência é a melhor escolha, quando comparada com Gabor, proporcionando uma melhor separabilidade do espaço. O índice da silhueta reforça esta percepção, conforme a projeção com características de co-ocorrência obtém 0,583, contra 0,474 para o experimento de Gabor, como visto na Figura 4.10.

A fim de identificar quais medidas estatísticas estão influenciando princi-

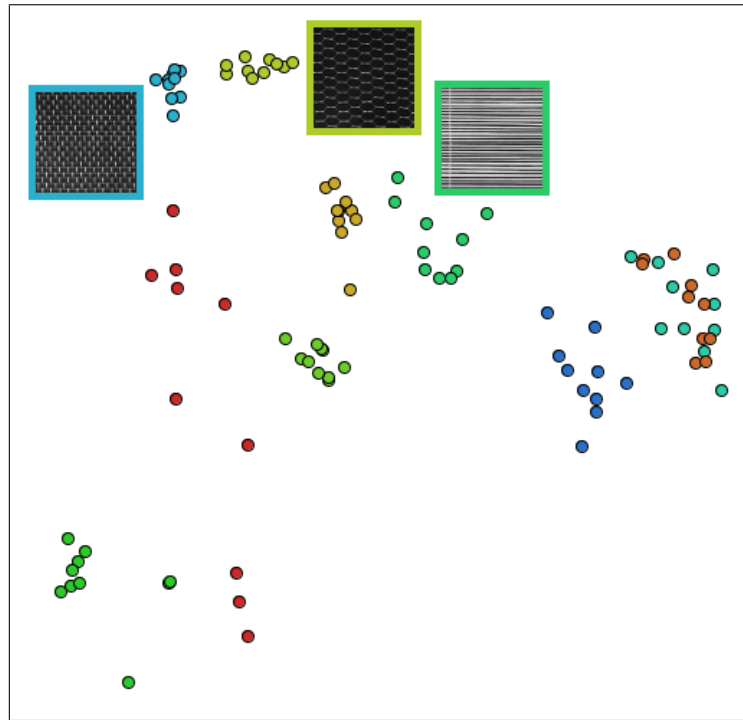


Figura 4.10: Características de Gabor usando apenas a orientação  $90^\circ$ . Duas das classes de imagens previamente adicionadas agora estão separadas, mas a coesão de todas as outras classes é reduzida; um resultado difícil de “ver” sem o apoio de visualização.

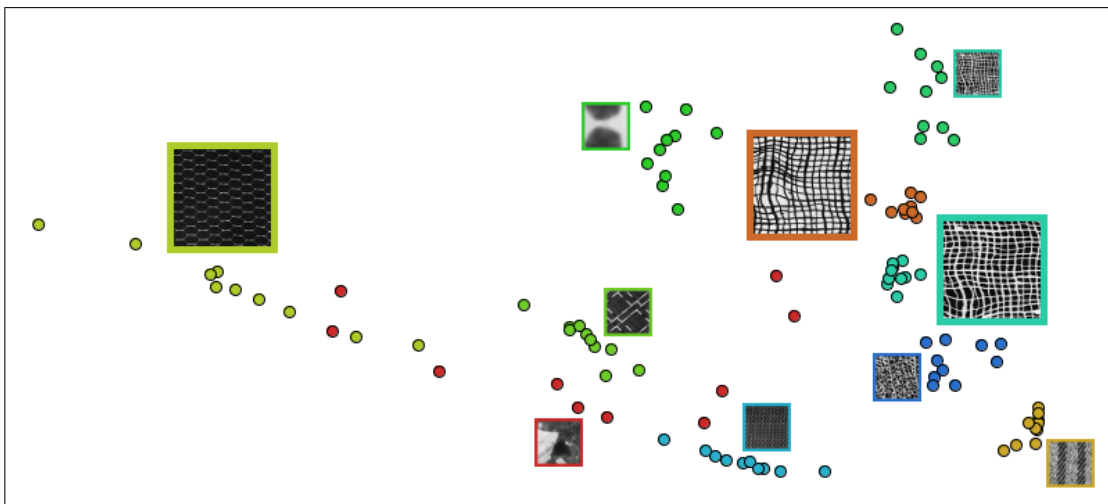


Figura 4.11: Projeção das classes selecionadas de *Broadt* usando características extraídas pela técnica matriz de co-ocorrência. As duas classes de linha, não devidamente separadas pelas características de Gabor, indicam que a matriz de co-ocorrência é mais adequada para este conjunto de classes.

palmente a dispersão desta classe de textura, foram criadas projeções considerando cada medida estatística separadamente. Medidas calculadas para os índices de silhuetas de tais projeções revelam que a medida de energia é a que produz os piores resultados, i.e, o menor valor de silhueta. Como consequência, uma nova projeção com todas as características, exceto energia, foi criada e mostrada na Figura 4.12. A classe arame se mostra mais bem

agrupada, enquanto que a coesão na formação de grupos das outras classes é preservada. Pode-se inferir que a energia é uma informação que compromete a representação do conjunto de dados sob análise. O índice da silhueta para essa projeção, 0,607, é maior do que aqueles apresentados nas configurações anteriores e indica que o conjunto de características usado aqui não somente é quantitativamente o melhor das representações, mas também coincide com a inspeção visual.

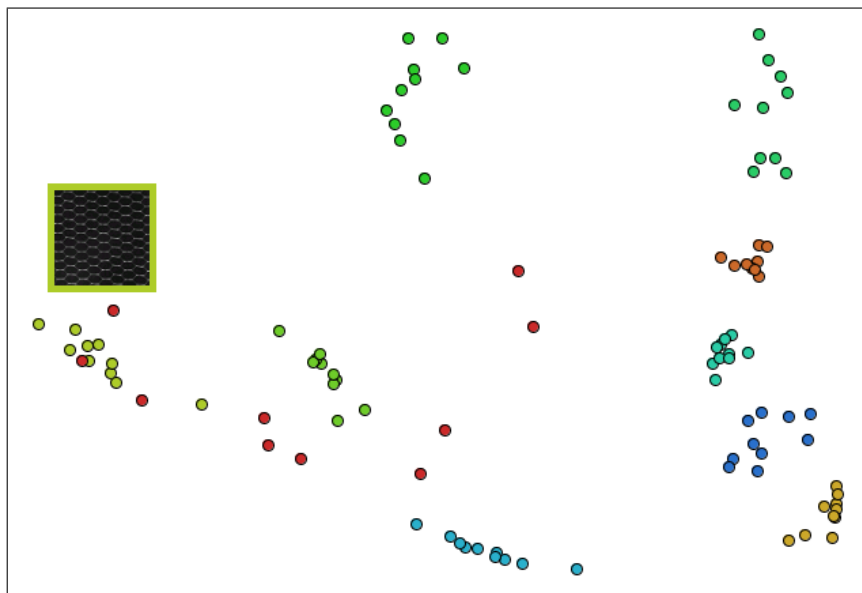


Figura 4.12: Projeção das texturas de *Brodatz* usando características computadas pela técnica matriz de co-ocorrência, sem a medida de energia. A projeção resultante apresenta tanto a melhor qualidade visual, quanto o maior valor do índice da silhueta. Isto indica que as características de energia são “ruidosas” para esse conjunto de dados.

## Experimento 2

Os exemplos a seguir buscam apresentar como a abordagem proposta se adapta à tarefa de classificação baseada em similaridade das amostras, mostrando que, conforme o índice da silhueta aumenta, as taxas de classificação também aumentam. Neste experimento, é usado um subconjunto de texturas do conjunto **Textured Surfaces**, contendo 280 imagens divididas em sete classes. A Tabela 4.1 mostra os resultados com o classificador K-NN, com média de classificação correta usando validação cruzada 10 dobras para características de Gabor computadas por diferentes orientações. Quatro testes, nomeados  $Teste_1$ ,  $Teste_2$ ,  $Teste_3$  e  $Teste_4$ , correspondem às orientações  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ , respectivamente. O  $Teste_5$  é a combinação das orientações  $0^\circ$  e  $90^\circ$ , enquanto que  $Teste_{Todas}$  consiste do teste com todas as orientações combinadas. Nota-se que os índices de silhuetas e as taxas de classificação são relacionados, corroborando a ideia de que se um usuário utilizar a melhor



projeção na escolha de um conjunto, os melhores resultados acompanharão a classificação do algoritmo K-NN.

Tabela 4.1: As quatro primeiras colunas da tabela mostram as ponderações para as características de Gabor. Cada parametrização gera uma projeção diferente, avaliadas pelo índice da silhueta. A última coluna sustenta a hipótese que os resultados visuais condizem com o desempenho de classificação.

$\theta$	$0^\circ$	$45^\circ$	$90^\circ$	$135^\circ$	Silhueta	Classificação (%)
$Teste_1$	1	0	0	0	0.1117	69.28
$Teste_2$	0	1	0	0	-0.0164	60.01
$Teste_3$	0	0	1	0	0.0164	68.57
$Teste_4$	0	0	0	1	-0.0179	53.93
$Teste_5$	1	0	1	0	0.2204	84.28
$Teste_{Todas}$	1	1	1	1	0.2758	87.85

No segundo teste são extraídas características de co-ocorrência do mesmo conjunto de imagens e combinadas com características de Gabor (orientações  $0^\circ$  e  $90^\circ$ ). As medidas estatísticas Entropia (H), Correlação (COR), Momento de Diferença Inversa (IDM) e Inércia (I) são computadas das matrizes de co-ocorrência. A Tabela 4.2 mostra os resultados do índice da silhueta e taxas de classificação. O  $Teste_1$  combina  $COR$  e  $IDM$ , enquanto  $Teste_{Todas}$  combina todas as características. Além disso, o  $Teste_{Melhores}$  representa as melhores características de acordo com os valores da silhueta das projeções, considerando cada característica individual. As características combinadas são filtros de Gabor com orientações  $0^\circ$  e  $90^\circ$ , juntamente com  $COR$  e  $I$  computadas da matriz de co-ocorrência. Novamente, é confirmada a ideia de que as melhores representações visuais (alcançando silhuetas maiores) convergem para taxas de classificação mais altas.

Tabela 4.2: As cinco primeiras linhas correspondem às combinações das características de Gabor e co-ocorrência. O índice da silhueta calculado pelas projeções resultantes e as taxas de classificação são mostradas nas últimas duas colunas, respectivamente.

Combinação	$0^\circ$	$90^\circ$	H	COR	IDM	I	Silhueta	Classificação (%)
$Teste_1$	0	0	0	1	1	0	0.2590	81.96
$Teste_{Todas}$	1	1	1	1	1	1	0.3023	84.18
$Teste_{Melhores}$	1	1	0	1	0	1	0.3899	84.61

Para avaliar a robustez desta abordagem em outros cenários, são criados conjuntos diferentes de características combinando Gabor e co-ocorrência. Para cada espaço mapeado na projeção é calculado o índice da silhueta e o mesmo subconjunto resultante é classificado. Os resultados deste experimento são apresentados na Figura 4.13. O eixo  $x$  indica os 24 conjuntos de características, e o eixo  $y$  mostra a medida de silhueta e as taxas de classificação obtidas a partir dos classificadores  $k$ -means (MacQueen (1967)), *Expectation*

*Maximization* (EM) (Mclachlan and Krishnan (2008)) e K-vizinhos mais próximos (K-NN). Para o classificador K-NN emprega-se  $K = 25$ . Para os outros, o número de *clusters* é definido como 7, cujo valor é a mesma quantidade de classes disponíveis. Pode-se observar no gráfico da Figura 4.13 que tanto as curvas do índice da silhueta como o desempenho de classificação apresentam a mesma tendência, isto é, quando a silhueta aumenta a classificação também aumenta, e vice-versa. Dessa maneira, pode-se inferir que a abordagem proposta pode ser um interessante guia para refinar interativamente o conjunto de características que representa adequadamente um conjunto de imagens.

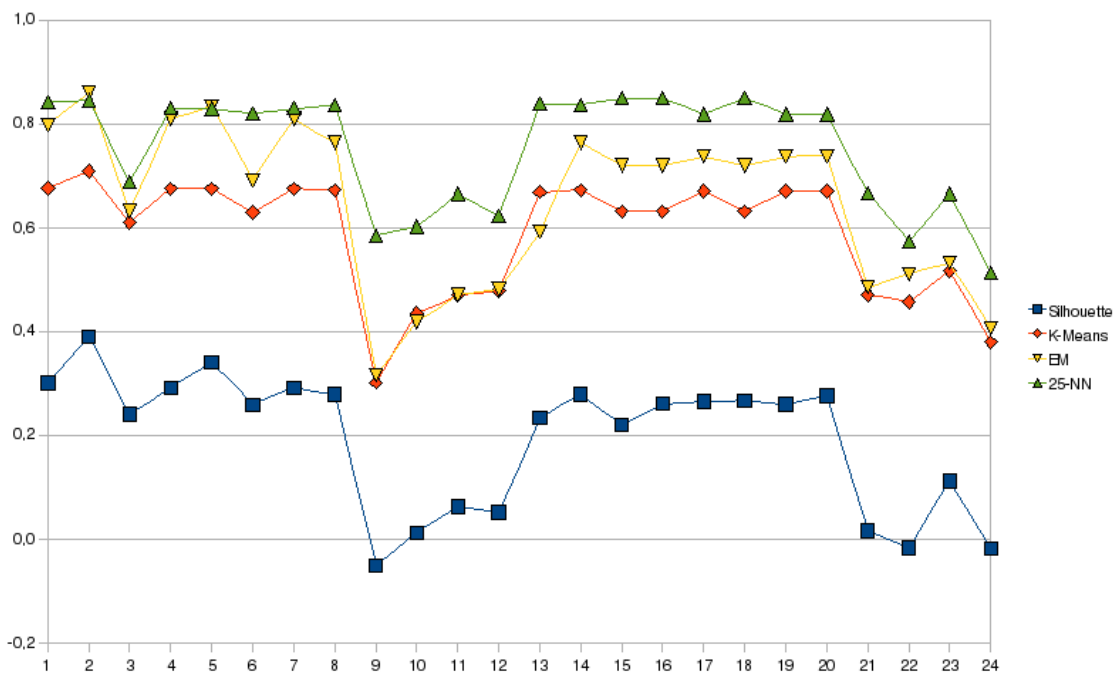


Figura 4.13: Relação entre o índice da silhueta e taxa de classificação para três diferentes classificadores. As curvas do índice da silhueta e o desempenho de classificação apresentam a mesma tendência, isto é, quando a silhueta aumenta a classificação aumenta, e vice-versa.

### Experimento 3

Este experimento avalia como o uso de técnicas de projeção, neste caso a técnica *Classical Scaling*, pode auxiliar especialistas a explorar o espaço de características construído por dados de treinamento. Dificilmente especialistas possuem a possibilidade de analisar o subconjunto de amostras usado para construir o modelo de classificação. Além disso, estas amostras devem representar ao máximo o conjunto inteiro, caso contrário, o risco de *overfitting* dos dados aumenta. Este problema tende a se agravar para dados com alta dimensionalidade. Esta ideia é usada para explorar o conjunto de imagens **ImageCLEF2006**, composto de 3.870 imagens divididas em 35 classes.

Aqui, são usadas 15 imagens de cada classe como conjunto de treinamento, produzindo um total de 525 imagens. A partir deste conjunto de dados, são extraídos três diferentes conjuntos de características: (1) 16 características de Gabor, (2) 100 características de co-ocorrência e (3) 1.024 características de transformadas *wavelets*. As projeções são criadas durante as etapas de teste e treinamento para cada conjunto de características. A Tabela 4.3 apresenta os resultados do índice de silhueta e de taxa de classificação usando o classificador K-NN, com  $K$  igual a 1 e 10. Os maiores valores para o índice da silhueta dos conjuntos de treinamento, também resultam em melhores taxas de classificação para os dados do conjunto inteiro. Essa correlação mostra que a qualidade objetiva das representações visuais de um subconjunto pode ser tomada como a interpretação da classificação para o conjunto de imagens inteiro, pois ela preserva o comportamento do modelo quando a seleção das imagens é representativa.

Tabela 4.3: Análise dos índices de silhueta e taxas de classificação para três diferentes conjuntos de características. Os resultados obtidos considerando o conjunto de treinamento é preservado quando o conjunto de imagens inteiro é considerado.

	Descritor	%(1-NN)	%(10-NN)	Silhueta
<b>Treinamento</b>	Gabor	18.86	16.95	-0.29
	Co-ocorrência	38.67	37.91	-0.22
	Wavelet	77.14	71.62	-0.19
<b>Teste</b>	Gabor	14.11	15.46	-0.34
	Co-ocorrência	35.16	36.08	-0.21
	Wavelet	77.10	75.67	-0.18

#### Experimento 4

Neste experimento é avaliada a abordagem proposta para a classificação de texturas de materiais do conjunto **KTH-TIPS**, comparando com os valores de silhueta obtidos das projeções. A Figura 4.14 mostra o desempenho de classificação para diferentes classificadores e os índices de silhueta das projeções resultantes. As características são computadas com a técnica de co-ocorrência para cinco valores distintos do parâmetro de distâncias (de 1 até 5). Além disso, é investigado o desempenho para todas combinações de características. O desempenho dos classificadores e os coeficientes de silhueta são indicados no eixo  $y$ , enquanto que as distâncias são indicadas no eixo  $x$ . Observe que as mudanças no índice da silhueta são refletidas no desempenho dos classificadores. Conforme o índice sobe ou desce, as taxas de classificação também crescem ou diminuem, confirmando a hipótese. A distância 1 oferece as melhores taxas de classificação, enquanto os valores que

variam entre 2 e 5, produzem taxas mais baixas. O mesmo ocorre com o índice da silhueta, condizendo com a hipótese. Esse resultado indica que as projeções e suas silhuetas podem ser usadas como um guia interessante para definir o parâmetro de distância do algoritmo de extração de características de co-ocorrência, segundo a validação do especialista.

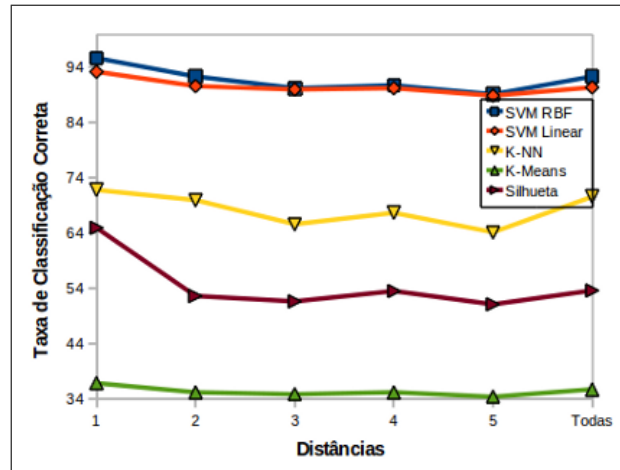
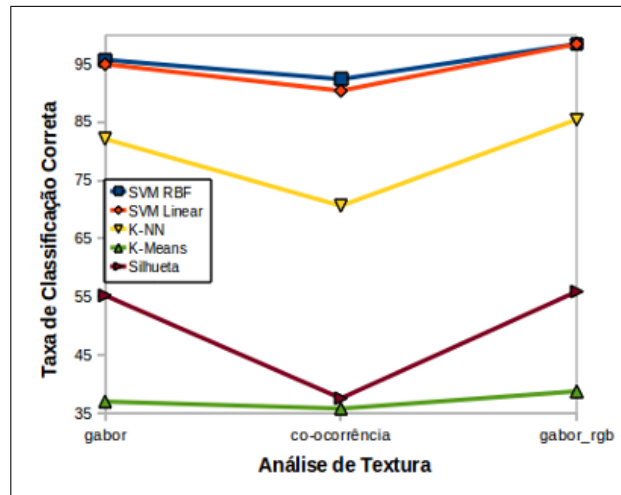


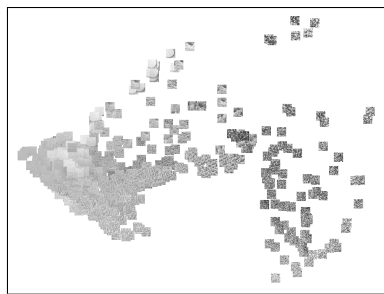
Figura 4.14: Avaliação do conjunto *KTH-TIPS* influenciado pelas distâncias da técnica matriz de co-ocorrência.

Ainda é avaliada a abordagem para técnicas tradicionais de análise de textura. A Figura 4.15 mostra os resultados de três descritores diferentes: filtros de Gabor (*gabor*), matriz de co-ocorrência (*co\_ocorrencia*) e filtros de Gabor aplicados sobre os canais R, G, B da imagem (*gabor\_rgb*). Observa-se que *gabor\_rgb* é superior tanto para a taxa de classificação correta, quanto para o índice da silhueta. Há uma clara distinção no desempenho entre *co\_ocorrencia* (*Todas* distâncias) e *gabor\_rgb*. A acurácia do classificador SVM com kernel RBF para *co\_ocorrencia*, *gabor* e *gabor\_rgb* é 90,37%, 94,93% e 98,39%, respectivamente. A silhueta também reflete essa acurácia, produzindo valores iguais a -0,3727, -0,2535 e -0,2505. As projeções são mostradas nas Figuras 4.15(b), 4.15(c) e 4.15(d). Analisa-se as projeções da esquerda para a direita e observa-se que as texturas mais finas são agrupadas neste mesmo sentido.

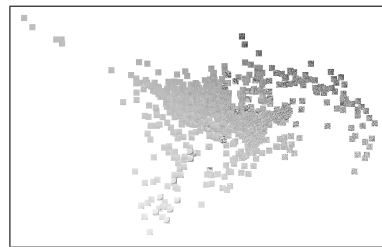
Com o objetivo de investigar a contribuição das abordagens de textura e cor para a separação deste conjunto, é empregado o modelo BoVF com amostragem densa com três descritores de cor: *histograma\_rgb*, *momentos\_invariantes* (Mindru et al. (2004)) e *sift* (Lowe (2004)). Na Figura 4.16(a), as taxas de classificação e os índices de silhueta são apresentados no eixo *y*, enquanto os descritores são apresentados no eixo *x*. Pode-se observar que *momentos\_invariantes* não executam bem. *histograma\_rgb* apresenta uma ligeira melhora, seguido pela técnica *sift*. Suas projeções são mostradas nas Figuras 4.16(b), 4.16(d) e 4.16(e), respectivamente. A Figura 4.16(c) mostra a mesma projeção da Figura 4.16(b), com círculos substituídos pelas imagens correspondentes. Para contornar a sobreposição visual, o centro da projeção é ampliado e mostrado na



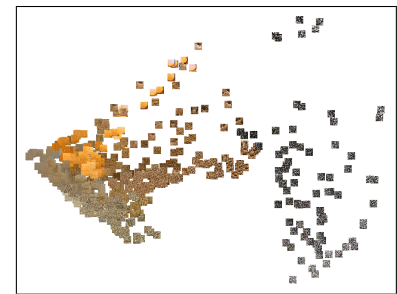
(a)



(b) Gabor



(c) Co-ocorrência



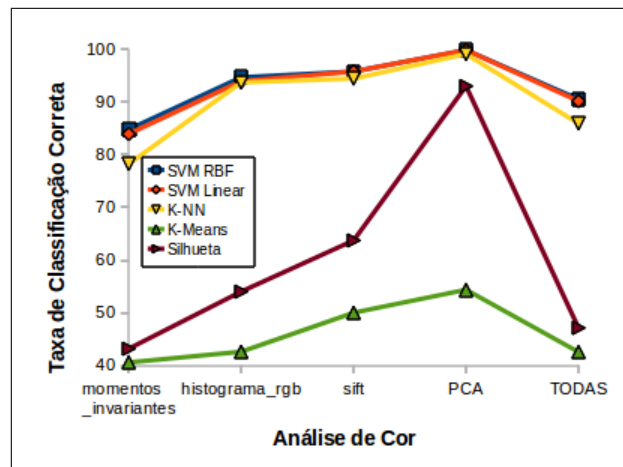
(d) RGBGabor

Figura 4.15: Comparação de diferentes conjuntos de características computados de técnicas tradicionais de análise de textura sobre o conjunto *KTH-TIPS*. Novamente as taxas de classificação correspondem aos índices de silhueta, mostrando a generalidade desta abordagem.

caixa no canto superior direito, da Figura 4.16(e). Embora haja uma mudança nos descritores, a hipótese é sustentada.

Normalmente, para conjuntos de imagens complexos, é desejável combinar abordagens diferentes de extração de características para capturar o máximo possível informações presentes nas imagens. Entretanto, os espaços de características resultantes podem ser de alta dimensionalidade, que na maioria dos casos, compromete a acurácia de classificação. Para questões de exploração, combinam-se os três descritores mencionados anteriormente e introduz-se uma etapa intermediária de redução de dimensionalidade por meio da *análise de componentes principais (PCA)* (Jolliffe (2002)), reduzindo o espaço para 10 dimensões. As taxas de classificação para o classificador K-NN são 85,92% e 99,12% para o espaço original e reduzido, condizendo com o índice da silhueta  $-0,2547$  e  $0,1290$ , respectivamente. As Figuras 4.16(f) e 4.16(g) apresentam a projeção para os descritores combinados e utilizando PCA. Finalmente, vale notar que o desempenho de classificação do K-NN é alto, se comparado ao

classificador SVM. Isso pode estar associado à boa separabilidade vista nas projeções das Figuras 4.16(d), 4.16(e) e 4.16(g).



(a)

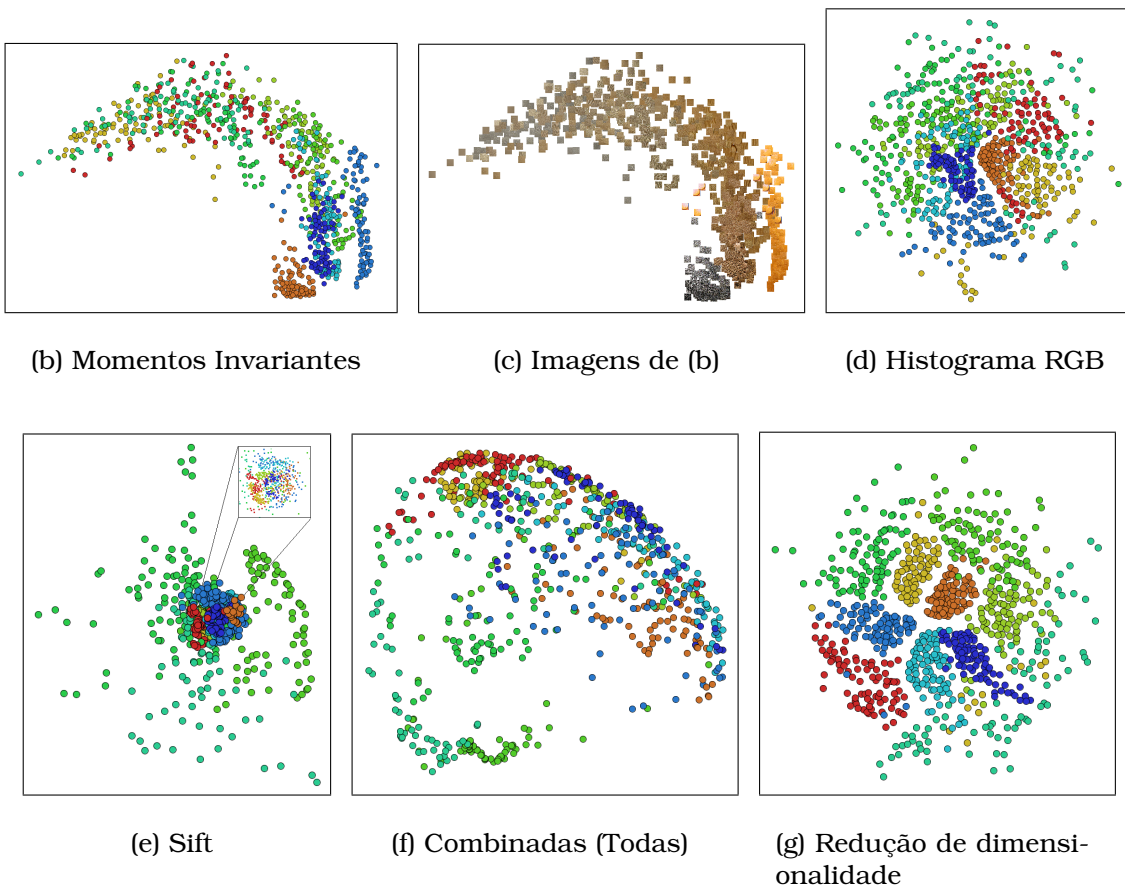
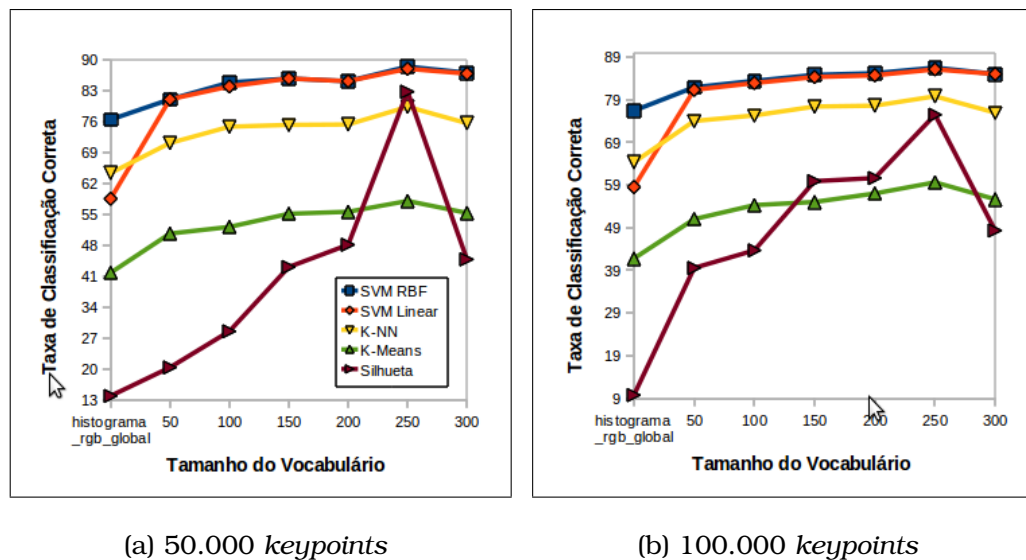


Figura 4.16: Projeções de espaços de características diferentes extraídas do conjunto de imagens *KTH-TIPS*. Apesar de PCA ter sido aplicado para reduzir a dimensionalidade, as melhores taxas de classificação e os valores de silhueta condizem com a separabilidade de classe observada na projeção.

## Experimento 5

No experimento **Corel** é apresentado um estudo comparativo da abordagem com o objetivo de avaliar as características locais e globais. O descritor histograma RGB é utilizado para ambas as abordagens. Características locais são computadas por detectores de região e quantizadas pelo modelo BoVF. O *histograma\_rgb* global é uma simples representação da distribuição dos pixels dos canais *R*, *G* e *B*. Portanto, histogramas de cor não consideram a informação espacial local, assim conjuntos de dados com fundos que apresentam cor similares o desempenho é realçado. A Figura 4.17 mostra as taxas de classificação no eixo *y*, enquanto as características globais e locais são indicadas no eixo *x*. Além disso, variam-se dois parâmetros do modelo BoVF: o tamanho do vocabulário e do número de *keypoints*. A Figura 4.17(a) e 4.17(b) mostram os resultados para o tamanho do vocabulário com 50.000 e 100.000 *keypoints*, respectivamente.



(a) 50.000 *keypoints*

(b) 100.000 *keypoints*

Figura 4.17: Resultados para tamanhos de vocabulários diferentes do modelo BoVF, considerando 50.000 e 100.000 *keypoints* amostrados sobre o conjunto de imagens *Corel*.

Pode-se observar na Figura 4.17 que a abordagem local apresenta desempenho significativamente melhor do que o global. A abordagem local é representada no gráfico pelos valores entre 50 e 300, correspondendo ao tamanho do vocabulário visual do modelo *bag-of-visual-features* (descrito na Seção 3.5). O maior coeficiente de silhueta ( $K = 250$ ) é 0,0055 para 50.000 *keypoints* e  $-0,0022$  para 100.000 *keypoints*. A silhueta para *histograma\_rgb* global é igual a  $-0,1029$ , o que representa uma redução significativa. O mesmo comportamento é observado para as taxas de classificação, que são 79,3% para 50.000 *keypoints* e 79,78% para 100.000, caindo para 64,30% no caso de *histograma\_rgb* global.

Com os melhores  $K$  em mãos, realizamos dois outros experimentos: 200.000 e 300.000 keypoints. O resultado mostra que a proposta é consistente, sendo um método útil para reduzir o tempo gasto na escolha do tamanho do vocabulário sobre o modelo BoVF.

### Experimento 6

Neste experimento, sobre o conjunto **Caltech**, novamente é avaliado o tamanho do vocabulário do modelo BoVF considerando as taxas de classificação e índices de silhueta. Entretanto, além do descritor *histograma\_rgb*, são usados dois diferentes descritores: *momentos\_invariantes* e *sift*. A Figura 4.18 mostra as taxas de classificação e índices de silhueta no eixo  $y$ , e o tamanho do vocabulário no eixo  $x$ . A Figura 4.18(a) e 4.18(b) mostram, respectivamente, os resultados para 50.000 e 100.000 keypoints usando *sift*. A taxa mais alta do coeficiente de silhueta e de classificação é obtida para  $K = 100$ .

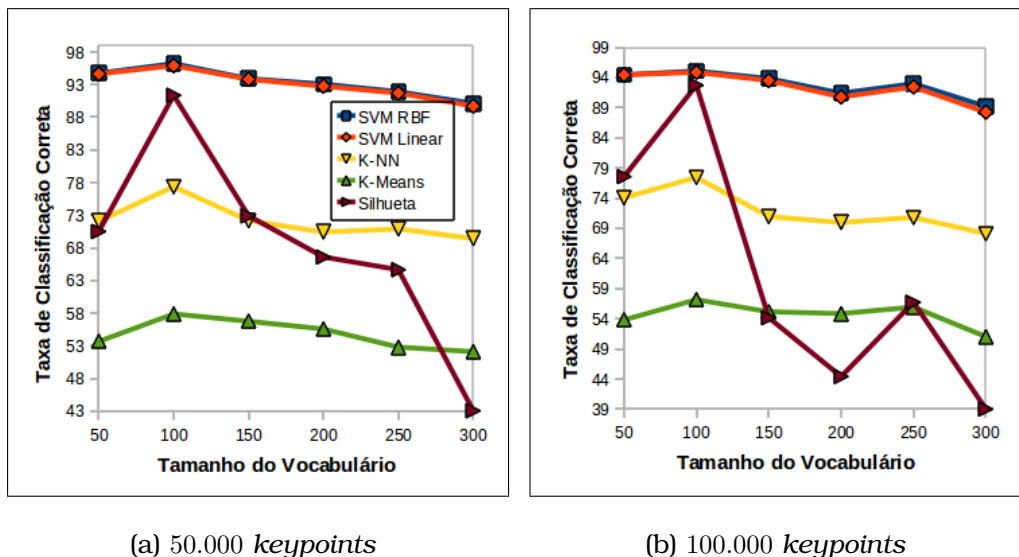


Figura 4.18: Resultados para diferentes tamanhos de vocabulário do modelo BoVF considerando 50.000 e 100.000 keypoints para o conjunto *Caltech*.

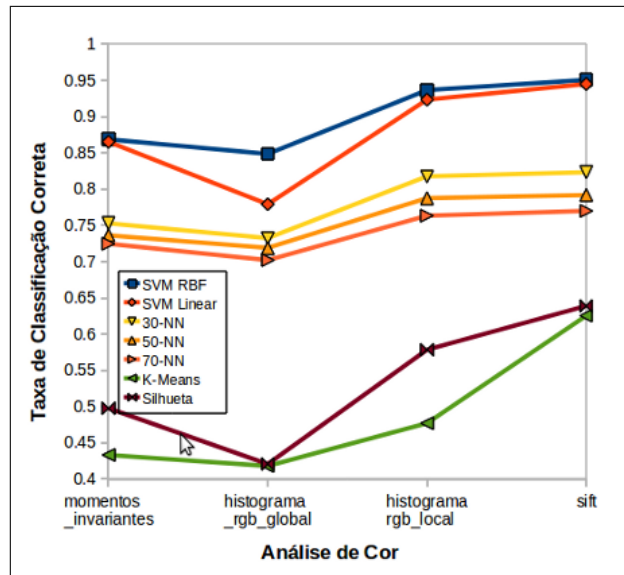
Esta proposta também explora características locais de cor usadas para construir o modelo BoVF, tendo  $K = 100$  como o melhor valor para o tamanho do vocabulário. A Figura 4.19, na página 58, mostra as taxas de classificação e índice da silhueta de três descritores: *momentos\_invariantes*, *histogramaRGB* e *sift*. Além disso, as características são extraídas por *histograma\_rgb* global. Pode-se ver que *histograma\_rgb* global obtém os piores resultados, enquanto que o descritor *momentos\_invariantes* é ligeiramente melhor. Para *histograma\_rgb* local, um aumento no desempenho é observado. Todas as taxas de classificação seguem o índice da silhueta obtido a partir da projeção dos conjuntos de características. O descritor *sift* exibe o melhor desempenho com 95,42%. Para



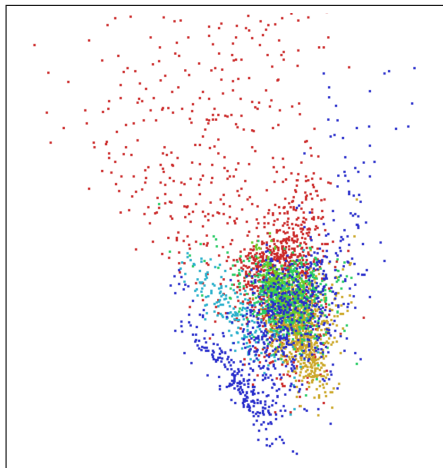
a classificação com o K-NN, são usados 30, 50 e 70 vizinhos. As Figuras 4.19(b) e 4.19(d) mostram as projeções para o melhor e o pior descritor, respectivamente. As projeções em que os círculos são substituídos por imagens são mostradas nas Figuras 4.19(c) e 4.19(e). É interessante notar que o método *histograma\_rgb* global é conhecido por sofrer forte influência da informação de fundo da imagem, que pode ser observado na projeção da Figura 4.19(c). Por outro lado, o resultado do descritor *sift* obtém uma projeção visualmente melhor, sendo fortemente recomendado para reconhecer as categorias de objeto, especialmente nos casos em que exista presença de variação intra-classe, e fundos diferentes sem a segmentação de objetos alvos. Assim, as projeções são ferramentas úteis para apoiar especialistas a tomar decisões sobre qual conjunto de características deve ser usado.

### 4.3 Considerações Finais

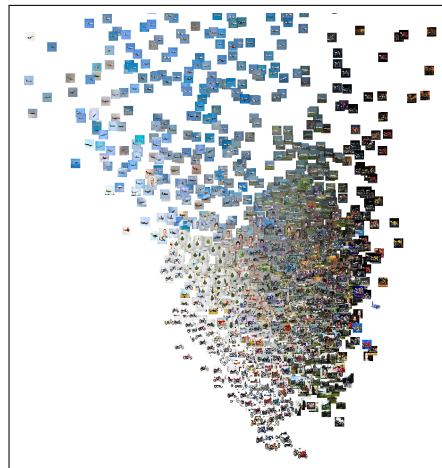
Este capítulo apresentou uma nova abordagem de exploração visual do espaço de características baseada em projeção multidimensional de dados. Com o objetivo de avaliar esta abordagem, experimentos foram realizados para seis conjuntos de imagens, descritos na Seção 4.2.1. O interesse aqui não está em alcançar o maior desempenho de classificação para os conjuntos de imagens analisados. Ao invés disso, foi mostrado que a *mineração visual* pode desempenhar um papel importante para entender o comportamento do espaço de características. Devido à falta de um protocolo experimental padrão, é difícil comparar a presente abordagem com outros resultados relatados na literatura, descritos no Capítulo 1. Os resultados deste capítulo são difíceis de comparar com resultados apresentados por Rodrigues Jr. et al. (2005), pois o banco de imagens médicas usado não foi disponibilizado publicamente. Podemos fazer comparações usando curva ROC sobre os conjuntos de imagens utilizado aqui. Porém este não é o foco do trabalho.



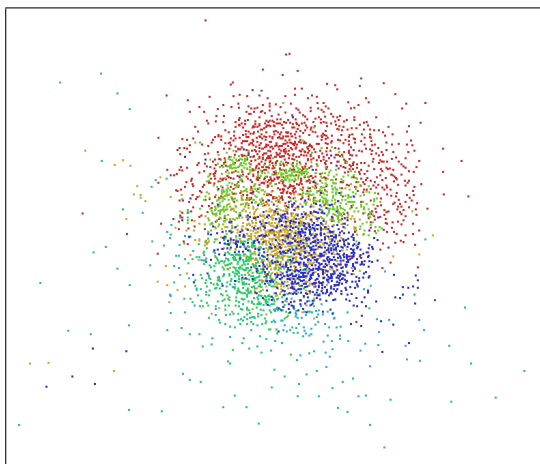
(a)



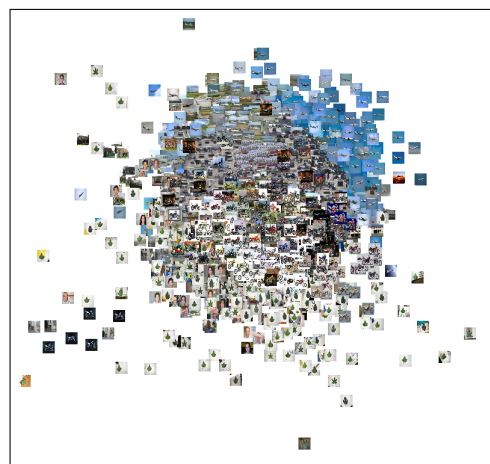
(b) Histograma RGB Global



(c) Imagens



(d) Sift



(e) Imagens

Figura 4.19: Comparação das projeções de dois espaços de características. A melhor projeção, em termos de separação de classes, coincide com as taxas de classificação, indicando que as projeções podem ser ferramentas úteis para apoiar os especialistas.

---

## CONCLUSÃO

---

Apesar de inúmeras propostas para descrição de imagens, e muitas delas publicadas recentemente, não há meios algorítmicos que sejam completamente aceitáveis para modelar a visão humana no contexto de interpretação de imagens. Isso recai em importantes aplicações de diversas áreas, incluindo *visão computacional*, *mineração de imagens* e *reconhecimento de padrões*. Assim, empregar técnicas de visualização de dados para explorar e extrair conhecimento de um conjunto de imagens é uma maneira eficiente de combinar a inteligência humana com a poderosa força da computação. Diversas técnicas e ferramentas têm sido propostas para permitir que usuários possam ser capazes de entender a representatividade dos dados, porém poucas delas projetadas para explorar conjuntos com espaços multidimensionais resultantes da etapa de descrição de imagens.

Este trabalho apresentou uma nova abordagem baseada em representações visuais para explorar espaços de características. Os resultados usando esta proposta confirmam a seguinte hipótese: conforme a qualidade das representações visuais melhora, o desempenho de classificação também melhora. Assim, as representações criadas por meio de técnicas de projeção podem ser empregadas com sucesso como um guia para entender o comportamento das características em termos de agrupamento de imagens similares e separação das dissimilares. Além disso, uma medida bastante aceita, o *índice da silhueta*, foi empregada para que usuários avaliassem a qualidade das projeções, superando problemas relacionados com a subjetividade de conclusões baseadas puramente em análise visual.

O presente trabalho não pretende ser uma abordagem para selecionar iterativamente o melhor conjunto de características. Ao invés disso, o foco está

em apoiar especialistas na difícil tarefa de entender o comportamento de uma técnica de descrição com diferentes vetores de características computados da variação de parâmetros. Além disso, a capacidade da exploração visual do conjunto de dados permite que usuários investiguem um dos maiores desafios em classificação de dados: a influência de classes que apresentam alta variação em suas amostras. A presença de *variação intra-classe* influencia diretamente a descrição das imagens, e em consequência, o desempenho de classificação. Dessa maneira, esta abordagem pode ser considerada como um processo realizado concomitante para a etapa de classificação, cuja ideia é auxiliar usuários, especialistas ou não, a identificar *outliers* e a rejeitar possibilidades de escolhas de algoritmos de descrição, acelerando o processo de busca pelo espaço de características que representa apropriadamente um conjunto de imagens.

## 5.1 Limitações

Embora seja demonstrada a utilidade desta abordagem para diversos conjuntos de imagens, uma limitação inerente às técnicas de projeção é a *escalabilidade visual*. Quando o número de instâncias de um conjunto de dados é grande, a representação visual deve ser capaz de sintetizar todos os pontos na tela evitando ao máximo a oclusão (Eick and Karr (2002)). Entretanto, há casos em que as técnicas de projeção criam distribuições de pontos não uniformes no espaço de visualização, sendo inevitável a sobreposição de pontos. De fato, a técnica escolhida para criar tais representações deve refletir as relações de distâncias presentes na estrutura dos dados originais. Isso revela ao especialista o quão adequadas as características são para representar um conjunto sob análise. Uma maneira de suavizar o problema de oclusão é aplicar o conceito de *representações contínuas* (van Liere and de Leeuw (2003)) para ganhar o conhecimento global da estrutura dos dados. Basicamente, os pontos de dados não são representados explicitamente, mas uma agregação de pontos que reflete a informação de densidade na visualização. A projeção final assemelha-se a um *mapa de calor* modelado por uma função gaussiana. Outra limitação, ainda pouco discutida na literatura, é como o posicionamento de *outliers* podem influenciar no valor médio do índice da silhueta.

## 5.2 Contribuições

As principais contribuições deste trabalho são:

- Uma abordagem para exploração visual de espaços de características

com o objetivo de convergir para características apropriadas na descrição de imagens, de acordo com o ponto de vista do especialista.

- Incorporação à ferramenta PEx-Image de uma estrutura de análise visual e interativa de dados que ajuda usuários a entender os diferentes conjuntos de características.
- Uma metodologia objetiva para avaliar a qualidade das projeções que condizem com o conceito de melhores projeções em espaços de características mais apropriados, especialmente para tarefas de classificação.

### 5.3 *Trabalhos Futuros*

Fica para desenvolvimentos futuros, a investigação de outras medidas objetivas que avaliam a qualidade das projeções. Entretanto, é importante que exista uma certa correlação entre os valores de tais medidas e a taxa de acerto de classificação, que permita usuários entenderem a adequabilidade do espaço de características. Além disso, esta abordagem pode ser facilmente estendida para outros tipos de aplicações, tais como recuperação multimídia. Muitas vezes, o acesso a arquivos de vídeo é difícil devido ao *gap semântico* entre o dado e a interpretação do usuário. Dessa maneira, representar objetos multimídia por vetores de características associado à interação visual pode tornar-se uma componente essencial em sistemas de recuperação baseada em conteúdo desta natureza.

Em trabalhos futuros pode ser estudada a inclusão de outros tipos de classificadores, e diferentes algoritmos de agrupamentos usados para analisar a informação estrutural dos dados, tais como *algoritmos hierárquicos* (Xu and Wunsch (2005)). Como mencionado anteriormente, a oclusão de pontos na representação visual geralmente ocorre em conjuntos com grande quantidade de instâncias. Uma resposta simples a este problema é utilizar *técnicas de interação* para explorar em maior nível de detalhes uma certa região. De outro ponto de vista, van Liere and de Leeuw (2003) propuseram uma técnica para fornecer a informação de densidade na estrutura da visualização. Embora a técnica *GraphSplatting* (van Liere and de Leeuw (2003)) suavize o problema de oclusão, ela não representa estruturas locais dos dados. Para este fim, é interessante criar uma *metáfora visual* que fosse complementar, isto é, que combine na mesma visualização ambas *visualização contínua e discreta*.

Recentemente, a descrição e o reconhecimento de *textura dinâmica* têm sido amplamente estudada pela comunidade de *visão computacional* (Fazekas et al. (2009); Traver et al. (2010)). Texturas dinâmicas são padrões visuais variantes no espaço e no tempo. Exemplos de aplicações incluem detecção

de fumaça em sistemas de segurança e reconhecimento de expressões faciais. Neste caso, as características estão relacionadas com os padrões de movimento. Como trabalhos futuros, a abordagem proposta aqui pode ser estendida para explorar visualmente a estrutura de texturas dinâmicas. Como este tipo de textura apresenta variação espaço-temporal, a sua representatividade pode ser analisada em visualizações de *séries temporais*. Além disso, uma sequência de projeções 2D pode ser exposta na tela com movimento contínuo das classes, resultando em uma animação de projeções. Isto pode responder questões relacionadas com o comportamento do algoritmo ao longo do tempo.

# Referências Bibliográficas

---

- Abdel-Hakim, A. E. and Farag, A. A. (2006). Csift: A sift descriptor with color invariant characteristics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR '06*, pages 1978–1983, Washington, DC, USA. IEEE Computer Society. Citado na página [29](#).
- Bianconi, F. and Fernández, A. (2007). Evaluation of the effects of gabor filter parameters on texture classification. *Pattern Recognition*, 40(12):3325–3335. Citado na página [22](#).
- Borg, I. and Groenen, P. J. (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer, second edition. Citado na página [10](#).
- Bosch, A., Zisserman, A., and Munoz, X. (2007). Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM International Conference on Image and Video Retrieval, CIVR '07*, pages 401–408, New York, NY, USA. ACM. Citado na página [29](#).
- Brodatz, P. (1966). *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York. Citado nas páginas [38](#) e [39](#).
- Burghouts, G. J. and Geusebroek, J.-M. (2009). Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113(1):48–62. Citado nas páginas [25](#), [29](#), [30](#), e [32](#).
- Chang, C.-C. and Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Citado na página [44](#).
- Chen, C. (2006). *Information Visualization: Beyond the Horizon*. Springer-Verlag New York, Inc., Secaucus, NJ, USA. Citado nas páginas [6](#) e [17](#).
- Cheng, H. D., Shan, J., Ju, W., Guo, Y., and Zhang, L. (2010). Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognition*, 43(1):299–317. Citado na página [20](#).

- Chi, Y. and Leung, M. K. H. (2008). A general shape context framework for object identification. *Computer Vision and Image Understanding*, 112(3):324–336. Citado na página [32](#).
- Cleveland, W. S. and McGill, R. (1985). Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833. Citado na página [5](#).
- Clough, P., Grubinger, M., Deselaers, T., Hanbury, A., and Müller, H. (2007). Overview of the image-clef 2006 photographic retrieval and object annotation tasks. In Peters, C., Clough, P., Gey, F. C., Karlgren, J., Magnini, B., Oard, D. W., de Rijke, M., and Stempfhuber, M., editors, *Proceedings of the 7th Cross-Language Evaluation Forum (CLEF) Conference on Evaluating Systems for Multilingual and Multimodal Information Access*, volume 4730 of *Lecture Notes in Computer Science*, pages 579–594, Berlin, Heidelberg. Springer-Verlag. Citado nas páginas [39](#) e [40](#).
- Corso, J. J. and Hager, G. D. (2009). Image description with features that summarize. *Computer Vision and Image Understanding*, 113:446–458. Citado nas páginas [29](#) e [30](#).
- Costa, L. d. F. and Cesar Jr., R. M. (2009). *Shape Classification and Analysis: Theory and Practice*. CRC Press, Inc., Boca Raton, FL, USA. Citado na página [32](#).
- Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27. Citado na página [44](#).
- Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling*. Chapman & Hall, London, UK, second edition. Citado nas páginas [3](#), [9](#), [10](#), [16](#), e [38](#).
- Cula, O. G. and Dana, K. J. (2001). Recognition methods for 3d textured surfaces. In Rogowitz, B. E. and Pappas, T. N., editors, *Proceedings of SPIE Conference Human Vision and Electronic Imaging VI*, volume 4299, pages 209–220, San Jose, CA, USA. SPIE. Citado na página [31](#).
- Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60. Citado nas páginas [20](#) e [32](#).
- de Leeuw, W. C. and van Liere, R. (2003). Visualization of multidimensional data using structure preserving projection methods. In *Data Visualization: The State of the Art*, pages 213–224, Boston, USA. Kluwer Academic Publishers. Citado na página [2](#).



- de Silva, V. and Tenenbaum, J. B. (2003). Global versus local methods in non-linear dimensionality reduction. In Becker, S., Thrun, S., and Obermayer, K., editors, *NIPS - Advances in Neural Information Processing Systems*, pages 705–712, Cambridge, MA, USA. MIT Press. Citado nas páginas 10, 13, e 17.
- Drimbarean, A. and Whelan, P. F. (2001). Experiments in colour texture analysis. *Pattern Recognition Letters*, 22(10):1161–1167. Citado nas páginas 21 e 23.
- Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218. Citado na página 11.
- Eick, S. G. and Karr, A. F. (2002). Visual scalability. *Journal of Computational and Graphical Statistics*, 11(1):22–43. Citado na página 62.
- Eler, D. M., Nakazaki, M. Y., Paulovich, F. V., Santos, D. P., Andery, G. F., Oliveira, M. C. F., Batista Neto, J., and Minghim, R. (2009). Visual analysis of image collections. *The Visual Computer: International Journal of Computer Graphics*, 25(10):923–937. Citado nas páginas 2, 15, e 17.
- Eler, D. M., Nakazaki, M. Y., Paulovich, F. V., Santos, D. P., Oliveira, M. C. F., Neto, J. E. S. B., and Minghim, R. (2008). Multidimensional visualization to support analysis of image collections. *Brazilian Symposium on Computer Graphics and Image Processing*, 0:289–296. Citado na página 15.
- Everitt, B. S., L. S. and Leese, M. (2001). *Cluster Analysis*. Arnold Publishers, London, UK. Citado na página 13.
- Faloutsos, C. and Lin, K.-I. (1995). Fastmap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *SIGMOD '95: Proceedings of the 1995 ACM SIGMOD international conference on Management of data*, pages 163–174, New York, NY, USA. ACM. Citado nas páginas 8 e 17.
- Fayyad, U., Grinstein, G. G., and Wierse, A. (2001). *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. Citado na página 15.
- Fazekas, S., Amiaz, T., Chetverikov, D., and Kiryati, N. (2009). Dynamic texture detection based on motion analysis. *International Journal of Computer Vision*, 82(1):48–63. Citado na página 63.
- Fergus, R., Perona, P., and Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 264–271, Los

- Alamitos, CA, USA. IEEE Computer Society. Citado nas páginas 31, 39, e 41.
- Finlayson, G. D., Hordley, S. D., and Xu, R. (2005). Convex programming colour constancy with a diagonal-offset model. In *ICIP - IEEE International Conference on Image Processing*, volume 3, pages 948–951. Citado na página 24.
- Forsyth, D. A. (1990). A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36. Citado nas páginas 23 e 24.
- Gabor, D. (1946). Theory of communication. *Journal of Institute of Electronic Engineering*, 93:429–457. Citado na página 22.
- Gevers, T. and Stokman, H. (2004). Robust histogram construction from color invariants for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):113–117. Citado na página 26.
- Gonzalez, R. C. and Woods, R. E. (2006). *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA. Citado nas páginas 23 e 25.
- Gool, L. J. V., Moons, T., and Ungureanu, D. (1996). Affine/photometric invariants for planar intensity patterns. In *ECCV: Proceedings of the 4th European Conference on Computer Vision*, pages 642–651, London, UK. Springer-Verlag. Citado na página 28.
- Grinstein, G., Buja, A., Asimov, D., and Inselberg, A. (1995). Visualizing multidimensional (multivariate) data and relations-perception vs. geometry. In *VIS '95: Proceedings of the 6th Conference on Visualization*, pages 405–411, Los Alamitos, CA, USA. IEEE Computer Society. Citado na página 5.
- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2002). Cluster validity methods: part i. *SIGMOD Record*, 31(2):40–45. Citado na página 14.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621. Citado na página 21.
- Harris, C. and Stephens, M. (1988). A combined corner and edge detection. In Mathews, M., editor, *Proceedings of The Fourth Alvey Vision Conference*, pages 147–151, Manchester, UK. Citado na página 31.
- Hayman, E., Caputo, B., Fritz, M., and Eklundh, J.-O. (2004). On the significance of real-world conditions for material classification. In *ECCV04 - 8th*

- European Conference on Computer Vision*, volume 4, pages 253–266. Citado nas páginas 31 e 39.
- Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8(2):179–187. Citado na página 27.
- Jan Flusser, T. S. and Zitová, B. (2009). *Moments and Moment Invariants in Pattern Recognition*. Wiley & Sons, New York, USA. Citado na página 28.
- Jia, Z., Amselag, L., and Gros, P. (2008). Content-based image retrieval from a large image database. *Pattern Recognition*, 41:1479–1495. Citado na página 30.
- Jolliffe, I. T. (2002). *Principal Component Analysis (second edition)*. Springer-Verlag, New York, NY, USA. Citado nas páginas 43 e 55.
- Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory*, IT-8:84–92. Citado nas páginas 20 e 30.
- Julesz, B. (1966). Perception. *Psychophys*, 1:300–302. Citado na página 20.
- Jurie, F. and Triggs, B. (2005). Creating efficient codebooks for visual recognition. In *Computer Vision, IEEE International Conference on*, pages 604–610, Los Alamitos, CA, USA. IEEE Computer Society. Citado nas páginas 30, 32, e 42.
- Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley Interscience, New York, USA. Citado nas páginas 13 e 14.
- Ke, Y. and Sukthankar, R. (2004). Pca-sift: A more distinctive representation for local image descriptors. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 506–513, Los Alamitos, CA, USA. IEEE Computer Society. Citado na página 29.
- Keim, D. A. (2000). Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics*, 6(1):59–78. Citado na página 15.
- Keim, D. A., Mansmann, F., Schneidewind, J., Thomas, J., and Ziegler, H. (2008). *Visual Analytics: Scope and Challenges*, pages 76–90. Springer. Lecture Notes in Computer Science (LNCS). Citado nas páginas 2 e 6.
- Kohonen, T. (1981). Automatic formation of topological maps of patterns in a self-organizing system. In Oja, E. and Simula, O., editors, *Proceedings of 2SCIA, Scandinavian Conference on Image Analysis*, pages 214–220, Helsinki, Finland. Citado nas páginas 7 e 17.

- Kohonen, T., Schroeder, M. R., and Huang, T. S., editors (2001). *Self-Organizing Maps*. Springer-Verlag New York, Inc., Secaucus, NJ, USA. Citado na página 7.
- Lazebnik, S. and Raginsky, M. (2009). Supervised learning of quantizer codebooks by information loss minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1294–1309. Citado nas páginas 30, 31, e 32.
- Lazebnik, S., Schmid, C., and Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1265–1278. Citado nas páginas 31, 38, e 39.
- Lazebnik, S., Schmid, C., and Ponce, J. (2006). 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*, CVPR '06, pages 2169–2178, Washington, DC, USA. IEEE Computer Society. Citado nas páginas 30 e 31.
- Leung, T. and Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43:29–44. Citado nas páginas 30 e 31.
- Li, J. and Wang, J. Z. (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088. Citado nas páginas xi, 6, 39, e 40.
- Li, J. X. (2004). Visualization of high-dimensional data with relational perspective map. *IEEE Information Visualization*, 3(1):49–59. Citado na página 17.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110. Citado nas páginas 29, 32, 42, 43, e 54.
- Lukac, R. and Plataniotis, K. N. (2007). *Color Image Processing: Methods and Applications*. CRC Press, Taylor-Francis Group, New York, USA. Citado nas páginas 26 e 27.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press. Citado nas páginas 32, 43, 44, e 50.
- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1995). *Multivariate Analysis (Probability and Mathematical Statistics)*. Academic Press. Citado na página 38.

- Martin-Merino, M. and Muñoz, A. (2004). A new sammon algorithm for sparse data visualization. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 1*, pages 477–481, Washington, DC, USA. IEEE Computer Society. Citado na página 7.
- Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). Robust wide baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767. Citado nas páginas 32 e 42.
- McLachlan, G. J. and Krishnan, T. (2008). *The EM Algorithm and Extensions*. Wiley-Interscience, Hoboken, USA, 2 edition. Citado na página 50.
- Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaf-falitzky, F., Kadir, T., and Gool, L. V. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65:43–72. Citado nas páginas 31, 32, 42, e 43.
- Milligan, G. and Cooper, M. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179. Citado na página 14.
- Mindru, F., Tuytelaars, T., Van Gool, L., and Moons, T. (2004). Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 94(1-3):3–27. Citado nas páginas 23, 25, 28, 43, e 54.
- Minghim, R., Paulovich, F., and Andrade Lopes, A. (2006). Content-based text mapping using multi-dimensional projections for exploration of document collections. In Erbacher RF, Roberts JC, G. M. B. K., editor, *Visualization and Data Analysis, SPIE-IS&T Electronic Imaging*, volume 6060, page 60600S. ACM. Citado nas páginas 8 e 17.
- Neisser, U. (1964). Visual search. *Scientific American*, 210(6):94–102. Citado na página 30.
- Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D., Yanker, P., Faloutsos, C., and Taubin, G. (1993). Qbic project: querying images by content, using color, texture, and shape. volume 1908, pages 173–187. SPIE. Citado na página 30.
- Ojala, T., Pietikäinen, M., and Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 24:971–987. Citado na página 31.

- Ommer, B. and Buhmann, J. M. (2010). Learning the compositional nature of visual object categories for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:501–516. Citado na página [30](#).
- Pampalk, E., Goebel, W., and Widmer, G. (2003). Visualizing changes in the structure of data for exploratory feature selection. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 157–166, New York, NY, USA. ACM. Citado na página [1](#).
- Paulovich, F., Oliveira, M., and Minghim, R. (2007). The projection explorer: A flexible tool for projection-based multidimensional visualization. In *SIBGRAPI '07: Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007)*, pages 27–34, Washington, DC, USA. IEEE Computer Society. Citado na página [15](#).
- Paulovich, F. V. (2008). *Técnicas geométricas para análise visual de dados: Integrando mineração e visualização*. PhD thesis, Universidade de São Paulo. Citado nas páginas [6](#) e [38](#).
- Paulovich, F. V. and Minghim, R. (2006). Text map explorer: a tool to create and explore document maps. In *IV '06: Proceedings of the conference on Information Visualization*, pages 245–251, Washington, DC, USA. IEEE Computer Society. Citado nas páginas [8](#) e [15](#).
- Paulovich, F. V., Nonato, L. G., and Minghim, R. (2006). Visual mapping of text collections through a fast high precision projection technique. *International Conference on Information Visualisation*, 0:282–290. Citado na página [15](#).
- Paulovich, F. V., Nonato, L. G., Minghim, R., and Levkowitz, H. (2008). Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics*, 14(3):564–575. Citado nas páginas [6](#), [8](#), e [17](#).
- Pontil, M. and Verri, A. (1998). Support vector machines for 3d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:637–646. Citado na página [30](#).
- Rodrigues Jr., J. F., Castañón, C. A. B., Traina, A. J. M., and Traina Jr., C. (2003). Using efficient visual exploration techniques to evaluate features for content-based image retrieval. In *XVI Brazilian Symposium on Computer Graphics and Image Processing*, pages 183–190, Los Alamitos, CA, USA. IEEE Computer Society. Citado na página [2](#).

- Rodrigues Jr., J. F., Traina, A. J. M., and Traina Jr., C. (2005). Enhanced visual evaluation of feature extractors for image mining. In *The 3rd ACS/IEEE International Conference on Computer Systems and Applications*, volume 1, pages 45–I, Washington, DC, USA. IEEE Computer Society. Citado nas páginas 2 e 57.
- Rothwell, C. (1994). Hierarchical object description using invariants. In *Proceedings of the Second Joint European - US Workshop on Applications of Invariance in Computer Vision*, pages 397–414, London, UK. Springer-Verlag. Citado na página 30.
- Rousseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(1):53–65. Citado na página 14.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Journal of Molecular Biology Evolution*, 4(4):406–25. Citado nas páginas 9 e 17.
- Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5):401–409. Citado nas páginas 7 e 17.
- Schiele, B. and Crowley, J. L. (2000). Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36:31–50. Citado na página 30.
- Schmid, C. and Mohr, R. (1997). Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:530–535. Citado na página 30.
- Schreck, T., Schneidewind, J., and Keim, D. (2008). An image-based approach to visual feature space analysis. In *WSCG 2008, 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pages 223–230, Plzen, Bory, Czech Republic. Citado nas páginas 1 e 2.
- Schreck, T., von Landesberger, T., and Bremm, S. (2010). Techniques for precision-based visual analysis of projected data. *Information Visualization*, 9(3):181–193. Citado nas páginas 1 e 14.
- Seo, J. and Shneiderman, B. (2005). A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization*, 4(2):96–113. Citado na página 17.

- Seo, J. and Shneiderman, B. (2006). Knowledge discovery in high-dimensional data: Case studies and a user survey for the rank-by-feature framework. *IEEE Transactions on Visualization and Computer Graphics*, 12:311–322. Citado na página 17.
- Seung, S. H. and Lee, D. D. (2000). The manifold ways of perception. *Science*, 290(12):2268–2269. Citado na página 9.
- Shafer, S. A. (1985). Using color to separate reflection components. *Journal of Color Research and Applications*, 10(4):210–218. Citado nas páginas 23 e 24.
- Sivic, J. and Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, volume 2 of ICCV '03, pages 1470–1477, Los Alamitos, CA, USA. IEEE Computer Society. Citado na página 30.
- Sonka, M., Hlavac, V., and Boyle, R. (2008). *Image Processing, Analysis, and Machine Vision*. Thomson, third edition. Citado na página 20.
- Swayne, D. F., Lang, D. T., Buja, A., and Cook, D. (2003). Ggobi: evolving from xgobi into an extensible framework for interactive data visualization. *Journal of Computational Statistics and Data Analysis*, 43(4):423–444. Citado na página 17.
- Taubin, G. and Cooper, D. B. (1992). Geometric invariance in computer vision. chapter Object recognition based on moment (or algebraic) invariants, pages 375–397. MIT Press, Cambridge, MA, USA. Citado na página 30.
- Tejada, E., Minghim, R., and Nonato, L. G. (2003). On improved projection techniques to support visual exploration of multidimensional data sets. *Information Visualization*, 2(4):218–231. Citado nas páginas 8, 17, e 37.
- Theodoridis, S. and Koutroumbas, K. (2006). *Pattern Recognition*. Academic Press. Citado na página 1.
- Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419. Citado nas páginas 11 e 17.
- Traver, V. J., Mirmehdi, M., Xie, X., and Montoliu, R. (2010). Fast dynamic texture detection. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *Proceedings of the 11th European Conference on Computer Vision (ECCV2010)*, pages 680–693, Heidelberg (Greece). Springer. Citado na página 63.



- Tuceryan, M. and Jain, A. K. (1998). Texture analysis. In *The Handbook of Pattern Recognition and Computer Vision (2nd Edition)*, pages 235–276. Citado na página 20.
- van de Weijer, J. (2005). *Color Features and Local Structure in Images*. PhD thesis, University of Amsterdam. Citado nas páginas 23 e 25.
- van de Weijer, J., Gevers, T., and Bagdanov, A. D. (2006). Boosting color saliency in image feature detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:150–156. Citado na página 26.
- van Liere, R. and de Leeuw, W. (2003). Graphsplatting: Visualizing graphs as continuous fields. *IEEE Transactions on Visualization and Computer Graphics*, 9(2):206–212. Citado nas páginas 62 e 63.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag, New York, NY, USA. Citado na página 44.
- Varma, M. and Garg, R. (2007). Locally invariant fractal features for statistical texture classification. In *ICCV International Conference on Computer Vision*, pages 1–8, Los Alamitos, CA, USA. IEEE Computer Society. Citado na página 31.
- Varma, M. and Zisserman, A. (2009). A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:2032–2047. Citado na página 31.
- Vendramin, L., Campello, R. J. G. B., and Hruschka, E. R. (2009). On the comparison of relative clustering validity criteria. In *Proceedings of the SIAM International Conference on Data Mining*, pages 733–744, Sparks, Nevada, USA. SIAM. Citado nas páginas 3 e 14.
- Wilkinson, L., Anand, A., and Grossman, R. (2006). High-dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1363–1372. Citado nas páginas 5, 14, e 17.
- Xu, R. and Wunsch, I. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678. Citado nas páginas 14 e 63.
- Young, G. and Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3(1):19–22. Citado na página 11.
- Zhang, J., Huang, H., and Wang, J. (2010). Manifold learning for visualizing and analyzing high-dimensional data. *IEEE Intelligent Systems*, 25:54–61. Citado nas páginas 9 e 10.

Zhu, L., Rao, A. B., and Zhang, A. (2002). Theory of keyblock-based image retrieval. *ACM Transactions on Information Systems*, 20:224–257. Citado na página [30](#).