
Reconhecimento de Doenças da Cultura da Soja
usando Vetores Localmente Agregados e
Atribuições Leves

Rillian Diello Lucas Pires
Orientador: Wesley Nunes Gonçalves

Rillian Diello Lucas Pires

Reconhecimento de Doenças da Cultura da Soja usando Vetores
Localmente Agregados e Atribuições Leves

Dissertação apresentada ao Programa de Mestrado *Stricto Sensu* em Ciência da Computação da Faculdade de Computação, mantido pela Universidade Federal do Mato Grosso do Sul, para o Exame de Dissertação, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação (Área de Concentração: Visão Computacional).

Orientador:
Prof. Dr. Wesley Nunes Gonçalves

Campo Grande
2017

Agradecimentos

- Agradeço a Deus pelo dom da vida e por me iluminar sempre e em todos os lugares, me dando a força para seguir em frente sempre pelo caminho mais correto;
- A minha família que me deu todo suporte sempre e foi minha fortaleza nos momentos de maior necessidade, em especial minha mãe, dona Zilda, e ao meu velho Vô Oswaldo que me cuida lá de cima;
- Ao meu orientador Prof. Dr. Wesley Nunes Gonçalves por ter me acompanhado até este momento e me guiado sempre com muita paciência e disponibilidade;
- A todos os meus amigos, em especial o pessoal do grupo Companheiros, 600ml e Santa Barbara FC.
- Aos meus companheiros de estudos Edilson, Jonhatan e Wesley Kanashiro que sempre me ajudaram quando necessário;
- A UFMS por ter disponibilizado espaço para a realização de estudos e demais atividades necessárias ao projeto;
- Aos meus professores de maneira geral, que me guiaram pelos caminhos do saber;
- A UFGD Dourados e ao projeto PlantVillage por terem fornecido as bases para experimentos e testes;
- A CAPES, pelo aporte financeiro para o projeto;
- E todos aqueles que contribuíram direta ou indiretamente com este trabalho.

Resumo

A agricultura brasileira vem crescendo, principalmente devido ao aumento da produtividade e o desenvolvimento de políticas de incentivo aos agricultores. Neste cenário, o carro chefe da produção agrícola brasileira são os grãos, em especial a soja. Apesar disso, as plantações brasileiras não tem atingido o potencial máximo de rendimento, e muito disso se deve a presença de doenças e pragas agrícolas que causam danos as plantas e conseqüentemente diminuem a produtividade. Desta forma, o uso de tecnologias e técnicas avançadas vem se tornando um diferencial, tanto é que atualmente diversos sistemas de computação são empregados na agropecuária para se obter melhores resultados. Nesse contexto, as técnicas baseadas em Visão Computacional tal como a classificação de imagens vem recebendo destaque, em especial a técnica do histograma de palavras visuais (do inglês *Bag-of-Visual-Words* - *BoVW*), que se baseia em construir um vocabulário de palavras visuais a partir de pontos de interesse detectados por descritores de pontos locais. Porém, apesar dos bons resultados obtidos pelo BoVW na classificação de imagens, ele utiliza uma forma de rotulação rígida que associa cada ponto de interesse a uma única palavra visual, o que gera perda de informações locais. Para contornar essa deficiência, foram propostos vetores de descritores localmente agregados (*Vector of Locally Aggregated Descriptors* - VLAD) que se baseiam na representação das palavras visuais por meio de um vetor de diferenças entre a palavra visual e os pontos descritos. Este trabalho tem como objetivo avaliar o BoVW e o VLAD na classificação de doenças foliares da cultura de soja e propor uma abordagem que permita a construção de descritores mais discriminativos através da atribuição leve, para tanto, foi construída uma aplicação que permite a utilização e avaliação de ambas as técnicas. Foram utilizados os descritores de pontos locais SIFT e SIFT Denso para detectar e descrever os pontos de interesse das imagens de folhas de soja. Para avaliar as técnicas propostas neste trabalho, foram realizados experimentos utilizando duas bases compostas por amostras de imagens de folhas de soja doentes e saudáveis, sendo que a primeira foi construída com o auxílio de um *scanner* e a segunda foi obtida através de uma câmera. Os resultados foram comparados com as técnicas existentes na literatura e, em ambas as bases os métodos propostos apresentaram resultados promissores. A inclusão de informações locais dos pontos, junto a atribuição leve dos descritores demonstrou percentuais de acerto acima de 90% em ambas as bases.

Palavras Chave: Descritores de Pontos Locais, Histograma de Palavras Visuais, Doenças Foliares da Cultura de Soja, Visão Computacional, Classificação de Imagens, Agricultura de Precisão.

Abstract

Brazilian agriculture has been growing, mainly due to increased productivity and the development of policies to encourage farmers. In this scenario, the flagship of Brazilian agricultural production are grains, especially soybeans. Despite this, Brazilian plantations have not reached the maximum yield potential, and much of this is due to the presence of diseases and agricultural pests that cause damage to the plants and consequently decrease productivity. In this way, the use of advanced technologies and techniques has become a differential, so much so that currently several computer systems are used in farming for better results. In this context, techniques based on Computer Vision such as image classification have been highlighted, especially the visual-word histogram (BoVW), which is based on Construct a vocabulary of visual words from points of interest detected by local point descriptors. However, despite the good results obtained by BoVW in image classification, it uses a form of rigid labeling that associates each point of interest with a single visual word, which generates loss of local information. In order to circumvent this deficiency, vectors of locally aggregated descriptors (VLAD) have been proposed. They are based on the representation of the visual words by means of a vector of differences between the visual word and the points described. This work aims to evaluate the BoVW and VLAD in the classification of foliar diseases of soybean crop and to propose an approach that allows the built of more discriminating descriptors through the light attribution, for that, an application has been developed that allows the use and Evaluation of both techniques. The SIFT and SIFT Denso local points descriptors were used to detect and describe the points of interest of soybean leaf images. To evaluate the techniques proposed in this work, experiments were carried out using two bases composed of samples of diseased and healthy soybean leaves, the first of which was constructed with the help of a scanner and the second was obtained through a camera. The results were compared with the existing techniques in the literature and in both bases the proposed methods presented promising results. The inclusion of local information of the points, along with the light attribution of the descriptors, showed success rates above 90% in both bases.

Keys Words: Local Points Descriptors, Histogram of Visual Words, Foliar Diseases of Soybean Culture, Computer Vision, Classification of Images, Precision Agriculture.

Sumário

Sumário	x
Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução	1
1.1 Motivação	1
1.2 Trabalhos Correlatos	5
1.3 Objetivos	7
1.3.1 Geral	7
1.3.2 Específicos	7
1.4 Estrutura do Texto	8
2 A soja	9
2.1 Doenças Foliares da Soja	10
2.1.1 Ferrugem Asiática (<i>Phakopsora Pachyrhizi</i>)	11
2.1.2 Mancha Alvo (<i>Corynespora Cassiicola</i>)	13
2.1.3 Míldio (<i>Peronospora Manshurica</i>)	13
2.1.4 Mancha Olho de Rã (<i>Cercospora Sojina</i>)	14
2.1.5 Mancha Parda (<i>Septoria Glycines</i>)	14
3 Referencial Teórico	16
3.1 Visão computacional e sua aplicação na agricultura	16
3.2 SIFT	18
3.2.1 Detecção de extremos	18
3.2.2 Localização e eliminação de pontos	20
3.2.3 Definição da orientação dos pontos	22
3.2.4 Construção do descritor	24
3.3 SIFT Denso	25
3.4 Histograma de Palavras Visuais (<i>Bag of Visual Words</i> - BoVW)	25
3.4.1 Detecção e Descrição de Pontos de Interesse	26
3.4.2 Criação do Vocabulário Visual	27
3.4.3 Construção de Histograma	28
3.5 Vetores de Descritores Localmente Agregados (<i>Vector of Locally Aggregated Descriptors</i> - VLAD)	29
3.5.1 Normalizações	32

4	Abordagens Propostas	33
4.1	Atribuição Leve e Pesos	35
4.2	VLAD-F: Vector of Locally Aggregated Descriptors - Full	36
4.3	VLAD-PNC: Vector of Locally Aggregated Descriptors P-Nearest Centroids	36
4.4	VLAD-VNNC: Vector of Locally Aggregated Descriptors - Variable Number of Nearest Centroids	37
5	Experimentos e Resultados	41
5.1	Base de Imagens de Soja	41
5.1.1	Soja Scanner	41
5.1.2	Plant Village	43
5.2	Resultados Experimentais	45
5.2.1	Resultados - Soja Scanner	46
5.2.2	Resultados - Plant Village	49
6	Conclusão	54
6.1	Trabalhos Futuros	55

Lista de Figuras

1.1	Evolução da produção das safras de soja nos últimos 5 anos [12].	2
1.2	Os pontos pequenos representam descritores extraídos da imagem e os círculos maiores são centroides. O triângulo representa um descritor, que é adequado a rotulação pois está bem próximo a um centroide. O quadrado representa um descritor ambíguo pois se encontra em uma região de fronteira e o losango pode representar um ruído.	4
2.1	Diagrama esquemático das inter-relações dos fatores envolvidos em epidemias de doenças de plantas. Adaptado de [25].	12
2.2	Sintomas da ferrugem asiática em folhas de soja. (A) ferrugem RB e em (B) ferrugem TAN. As lesões RB são caracterizadas por uma coloração marrom avermelhadas e são consideradas um tipo de lesão mais resistente quando comparadas com as lesões TAN.	12
2.3	Sintomas da Mancha alvo em folhas de soja. Fonte http://dirceugassen.com/	13
2.4	Sintomas de míldio em folhas de soja.	14
2.5	Sintomas da mancha-olho-de-rã em folhas de soja. Fonte https://www.agrolink.com.br/	15
2.6	Sintomas da Mancha Parda em folhas de soja. Fonte https://www.agrolink.com.br/	15
3.1	(a) Imagem original em escala de cinza. Em (b), (c) e (d) observa-se a convolução de (a) com núcleos Gaussianos com diferentes valores de σ . É possível observar que a medida que σ aumenta, a imagem fica mais suavizada. Finalmente, (e) e (f) apresentam as diferenças de Gaussianas.	20
3.2	Representação do procedimento de obtenção das Diferenças de Gaussianas DoG para diversas escalas de uma imagem. Imagem baseada em [40] e adaptada utilizando imagens de doenças foliares da soja.	21
3.3	Detecção de extremos no espaço-escala σ . Imagem obtida e adaptada de [40].	21
3.4	(a) Pontos obtidos a partir do DoG, (b) Pontos restantes após a eliminação de pontos com baixo contraste, (c) Pontos restantes após a eliminação de pontos em bordas. Imagem obtida e adaptada de [19].	23
3.5	(a) Região delimitada por uma função de peso Gaussiano, com círculo. Imagem adaptada de [40]. (b) Representação de um possível histograma utilizado para determinar orientação de um ponto de interesse. Imagem adaptada de [60].	23

3.6	Setas em amarelo indicam a orientação do ponto de interesse. Imagem adaptada de [19].	24
3.7	Representação da janela de gradientes e do descritor de ponto de interesse obtido através do SIFT. Imagem adaptada de [24].	25
3.8	Exemplo de grade utilizada no SIFT Denso. Cada círculo representa um ponto de interesse na grade densa.	26
3.9	Passos do algoritmo BoVW (a) Corresponde a detecção e descrição de pontos de interesse. (b) Corresponde a criação do vocabulário visual. (c) Apresenta a contagem de palavras visuais. (d) Corresponde a criação do histograma. Imagem adaptada de [59].	27
3.10	Descritores obtidos a partir do SIFT.	28
3.11	Construção do vocabulário visual a partir do <i>K-means</i> . (a) Conjunto de descritores extraídos, (b) Construção do vocabulário visual com o <i>K-means</i> , (c) Palavras visuais que compõem o vocabulário.	29
3.12	Construção do histograma da imagem, (a) Pontos de interesse rotulados pela palavra visual com menor distância Euclidiana. (b) Construção do histograma que representa o descritor final da imagem.	29
3.13	Rotulação dos descritores pela palavra visual mais próxima.	30
3.14	Cálculo da diferença entre $r_i - c_{r_i}$ para cada descritor rotulado por uma palavra visual.	31
3.15	Somatório das diferenças de cada descritor r_i ao centroide c_{r_i} correspondente, para a construção do vetor v final.	31
4.1	Etapas seguidas nas abordagens propostas para o cálculo do peso de um descritor para os centroides. (a) Obtenção dos descritores e centroides; (b) Cálculo da distância do descritor r_1 para os centroides, (c) O cálculo do peso do descritor r_1 (d) Aplicação dos pesos aos vetores VLAD.	34
4.2	Primeira abordagem: Todos os descritores são atribuídos a todos os centroides.	37
4.3	Segunda abordagem: atribui um descritor a um conjunto P de centroides com maior peso. Nessa imagem temos $P = 2$, assim cada descritor é rotulado a dois centroides.	38
4.4	Terceira abordagem: os descritores são atribuído a um número variável de centroides de acordo com o peso que eles tem para o centroide.	40
5.1	Processo de realização dos experimentos. 1) Fonte de dados, 2) Pré-Processamento feito com o auxílio de um especialista da área, 3) Construção de um banco de imagens com amostras das folhas doentes e saudáveis, 4) Aplicação de um dos descritores de pontos locais estudados (SIFT, SIFT Denso), 5) Construção do vocabulário de palavras visuais a partir dos descritores da imagens, 6) Processo de atribuição dos descritores as palavras visuais de acordo com a técnica selecionada e 7) Classificação por meio do SVM Linear.	42
5.2	Procedimento de obtenção das imagens utilizado nesse estudo. O banco é composto por 4 classes de imagens que representam o problema de classificação de doenças foliares da cultura de soja. Adaptada de [53].	42
5.3	Conjunto de imagens das classes do problema de identificação de doenças foliares da cultura de soja. (a) Classe de folhas com a doença ferrugem asiática RB, (b) Classe de folhas com a doença da ferrugem asiática TAN, (c) Classe de folhas com a doença míldio e (d) Classe de folhas saudáveis.	43

5.4	Conjunto de imagens da base Plant Village sem nenhum pré-processamento, disponível em www.plantvillage.org . Em (a) folha com a doença míldio, (b) folha com a doença mancha olho de rã, (c) folha com a doença da mancha parda, e (d) folha saudável.	44
5.5	Conjunto de imagens das classes do problema de identificação de doenças foliares da cultura de soja. (a) Classe de folhas com a doença míldio, (b) Classe de folhas com a doença mancha olho de rã, (c) Classe de folhas com a doença da mancha parda, e (d) Classe de folhas saudáveis.	45
5.6	Taxa de classificação correta na base de imagens Soja Scanner, obtida pelas técnicas propostas utilizando os descritores SIFT e SIFT Denso.	48
5.7	Matrizes de confusão que ilustram resultados obtidos na tarefa de classificação de doenças foliares junto a base de dados Soja Scanner 5.1.1.	49
5.8	Taxa de classificação correta na base de imagens Plant Village, obtida pelas técnicas propostas utilizando os descritores SIFT e SIFT Denso.	52
5.9	Matrizes de confusão que ilustram resultados obtidos na tarefa de classificação de doenças foliares junto a base de dados Plant Village 5.1.2.	53

Lista de Tabelas

1.1	Trabalhos que promoveram extensões nas técnicas de classificação de imagens a partir do BoVW.	5
2.1	Descrição dos estágios vegetativos da soja.	10
2.2	Descrição dos estágios reprodutivos da soja.	11
5.1	Comparação do desempenho das técnicas, em relação ao tamanho dos Vocabulários no banco de imagens Soja Scanner, utilizando o SIFT e o SIFT Denso como descritores de pontos locais.	47
5.2	Comparação do desempenho das técnicas, em relação ao tamanho dos Vocabulários no banco de imagens Plant Village, utilizando o SIFT e o SIFT Denso como descritores de pontos locais.	51

Introdução

1.1 Motivação

A agricultura brasileira tem crescido fortemente ancorada em aumentos crescentes de produtividade, em políticas públicas adequadas e no empreendedorismo do produtor rural. A produção brasileira de grãos mais que dobrou em duas décadas segundo estimativas da CONAB [12], principalmente em decorrência de ganhos de eficiência, sendo que nesse período a taxa de crescimento da produtividade (3,2%), foi quase duas vezes superior à da área plantada (1,7%).

Segundo Ministério da Agricultura [1], na safra 2013/14, apesar das adversidades climáticas que atingiram importantes regiões de produção agrícola, o Brasil colheu aproximadamente 195,47 milhões de toneladas de grãos. Para a safra 2014/15 foram colhidas aproximadamente 209,5 milhões de toneladas. Na para a safra 2015/2016 foram colhidas aproximadamente 186,4 milhões de toneladas. Redução de aproximadamente 10% em relação à safra 2014/15. Porém, a safra 2015/2016 trouxe um novo recorde, em que foram colhidas aproximadamente 234,3 milhões de toneladas de grãos. Com destaque para a produção de Soja que atingiu uma produção de 113.9 milhões de toneladas.

No cenário nacional, de acordo com CONAB [12], o Centro-Oeste brasileiro é o principal produtor da cultura de soja. Sendo que para a safra de 2016/2017 estima-se uma produção de 50,1 milhões de toneladas, com uma produtividade média de 3.301 kg por hectare. A segunda grande região produtora de soja no Brasil é a região sul, que para a safra 2016/2017 estima-se 40,5 milhões de toneladas e uma produtividade média de 3.537 kg por hectare. Os três maiores produtores estaduais de soja no Brasil são os Estados do Mato Grosso (30,5 milhões de toneladas), Paraná (19,5 milhões de toneladas) e Rio Grande do Sul (18,7 milhões de toneladas); Mato Grosso do Sul aparece em quinto com uma produção 8,5 milhões de toneladas e uma produtividade média de 3.400 kg por hectare.

O gráfico da Figura 1.1 ilustra o crescimento da produção de soja para todas as regiões brasileiras nos últimos 5 anos, segundo dados da CONAB [12]. Se pode evidenciar as regiões Centro-Oeste e Sul como maiores produtoras, e que nos últimos 5 anos apresentaram um crescimento estável. Nas demais regiões se pode destacar a notável queda de produção na região Nordeste durante a safra de 2015/2016, que fez com que a produção da região fosse

superada pela produção da região Sudeste nessa safra. A safra 2015/2016 foi marcada por diversos problemas relacionados ao clima e as chuvas que afetaram negativamente a produção nacional.

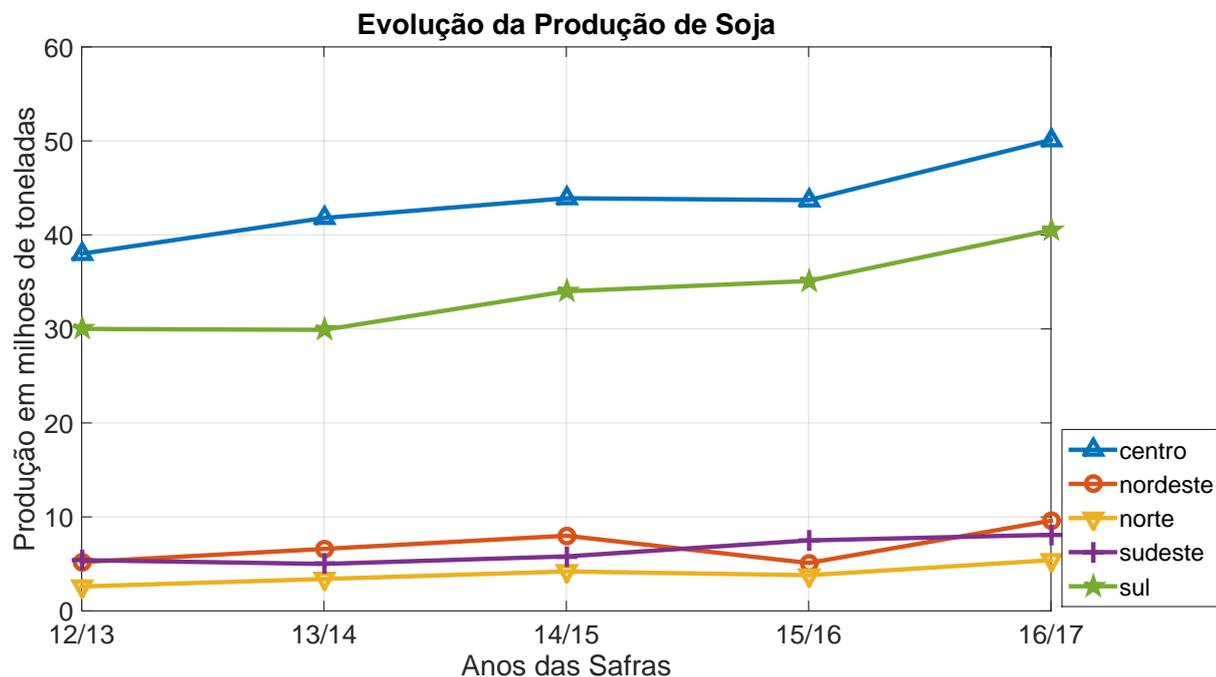


Figura 1.1: Evolução da produção das safras de soja nos últimos 5 anos [12].

Apesar da soja ser atualmente uma das principais culturas em nosso país, dada a sua expansão por todo o território nacional, a exploração econômica do seu potencial de rendimento (superior a 4.000 ha-1) dificilmente é alcançada por falta de manejo adequado [10]. Por exemplo a produção média por hectare na região Centro-Oeste é 3.301 kg e na região Sul é de 3.537 kg. A Organização Mundial para Alimentação e Agricultura (FAO) considera que as pragas¹ são responsáveis, em média, por cerca de 42,1% das perdas na produção. Segundo Kreyci et al [35] em 2013, se estimou que os fitopatógenos² foram responsáveis por 13,3% de danos. As plantas daninhas por 13,2%; insetos e ácaros por 15,6%.

Com a expansão da cultura de soja em novas áreas para o cultivo comercial, o número de doenças causadas por bactérias, fungos, nematoides e vírus, tende a aumentar limitando a produção e o rendimento [67] a cada safra. Além disso, a competitividade, exigências do mercado consumidor e a busca pela conservação dos recursos naturais através da diminuição da poluição do meio ambiente, são fatores que tem motivado a procura de um aprimoramento do manejo e gerenciamento dos processos produtivos agrícolas.

Uma das alternativas está no monitoramento das doenças desde a fase da semeadura até a colheita, sendo possível, quantificar os danos, diminuir os custos da aplicação de defensivos agrícolas³, e a diminuição do impacto ambiental; e ainda obter o aumento da produtividade e rendimento. Diante desse cenário, cresceu a importância da agricultura de precisão, que é definida por Pires *et al* [52] como "o uso de tecnologias para o manejo de solo, insumos e culturas, de modo adequado às variações espaciais e temporais em fatores que afetam a produtividade das mesmas com o objetivo de aumentar a eficiência produtiva".

¹ervas ou outra planta daninha, ou qualquer forma de vida animal que possa destruir aquilo que o ser humano considera um bem [29]

²agentes causadores de doenças

³Produto químico utilizado no combate e prevenção de pragas agrícolas, agrotóxicos

Atualmente, muitos sistemas são aplicados na área do agronegócio para obter melhores resultados no aumento da produção, qualidade dos produtos, redução de desperdícios e, ainda, aqueles que auxiliam na tomada de decisão. Dentre as ferramentas utilizadas na agricultura de precisão destaca-se o sensoriamento remoto, muito utilizado para o mapeamento de áreas de cultura e estimação do rendimento das mesmas. Entretanto, nos últimos anos a pesquisa tem sido direcionada para o monitoramento das condições da cultura como, por exemplo, a identificação das deficiências nutricionais das plantas, e mais recentemente a ocorrência de doenças, e classificação de plantas invasoras, dentre outros [23].

Nesse contexto, a visão computacional é uma área de pesquisa que vem avançando e auxiliando na evolução e desempenho dos processos realizados manualmente [7]. A visão computacional é uma área de pesquisa recente e responsável pela "visão" de uma máquina, extraindo informações significativas, possibilitando reconhecer, manipular e analisar os objetos que compõem uma determinada imagem. Segundo Prince [54] o principal objetivo da visão computacional é extrair informação útil das imagens. Essa tarefa tem se demonstrado surpreendentemente desafiadora, tendo despertado grande interesse dos pesquisadores nas últimas décadas.

Diversos projetos que utilizam técnicas de visão computacional para a agricultura de precisão vem sendo desenvolvidos, como por exemplo a detecção de danos foliares através de imagens digitais [17], identificação de plantas invasoras [55], identificação e contagem de leveduras [59], entre outros. Dentre os diversos problemas estudados na visão computacional que tem aplicações na agricultura de precisão, a classificação de imagens é um dos mais estudados dado o seu elevado grau de complexidade. Segundo Csurka et al [15], a classificação de imagens tem como objetivo identificar a presença de objetos em imagens, ou reconhecê-las como um tipo de cena em particular. No reconhecimento de doenças foliares, o objetivo é determinar a presença de características visuais nas folhas que são atribuídas as doenças, tais como manchas, rachaduras e ranhuras, como podemos observar na Figura 2.2 que apresenta a ferrugem asiática (*Phakopsora pachyrhizi*).

Atualmente a grande maioria dos métodos para a classificação de imagens utilizam a técnica do histograma de palavras visuais (do inglês *Bag-of-Visual-Words* - *BoVW*) proposta por Csurka et al [15]. O primeiro passo do BoVW consiste na detecção dos pontos de interesse. Existem hoje na literatura diversos algoritmos que tem como objetivo detectar e/ou descrever pontos. Destes, podemos citar o SIFT (*Scale Invariant Feature Transform*) [40], SIFT Denso [38], SURF (*Speeded-UP Robust Features*) [5], PHOW (*Pyramid Histogram Of Visual Words*) [8], HOG (*Histograms of Oriented Gradients*) [16] e o LBP (*Local Binary Patterns*) [44].

Após detectar e descrever o conjunto de pontos de interesse, o segundo passo compreende o agrupamento dos descritores. Usualmente o método utilizado para o agrupamento é o *K-means*, que segundo Fonseca [22], tem se mostrado uma alternativa relevante para realizar o agrupamento de palavras visuais por ser simples, intuitivo, e possuir uma complexidade linear. O *K-means* extrai um conjunto de centroides que constituem o vocabulário de palavras visuais. Com o vocabulário construído, cada ponto de interesse localizado na imagem é então associado à palavra visual mais próxima de acordo com a distância Euclidiana, em um processo conhecido como rotulação; e por fim é construído um histograma baseado na frequência de cada palavra visual na imagem que compõe o vetor de características da imagem.

Apesar dos bons resultados obtidos na classificação de imagens com o BoVW, ele utiliza a atribuição rígida [49], em que um descritor é associado a uma única palavra visual, incrementando a contagem da palavra visual no histograma. Essa forma de rotulação pode gerar perda de informações locais no espaço de características pois não leva em consideração se a localização do descritor esta em uma região de fronteira, se é um ruído, etc. Conforme Figura 1.2, os pequenos pontos representam descritores que foram extraídos, os círculos são centroides

obtidos por agrupamento. Na imagem foram selecionados 3 descritores, sendo representados por um triângulo, um quadrado e um losango. Utilizando-se a atribuição rígida nesse três descritores: o triângulo seria adequadamente rotulado pois está próximo da palavra visual e não tem relação com nenhuma outra. Já o quadrado está localizado em uma região de fronteira entre duas palavras visuais e sua rotulação rígida o associaria a apenas uma palavra visual sendo que ele tem influência sobre ambas. Por fim, o losango, seria associado a palavra visual mais próxima porém observamos que ele não se encontra de fato próximo a nenhuma palavra visual, o que caracteriza um ruído.

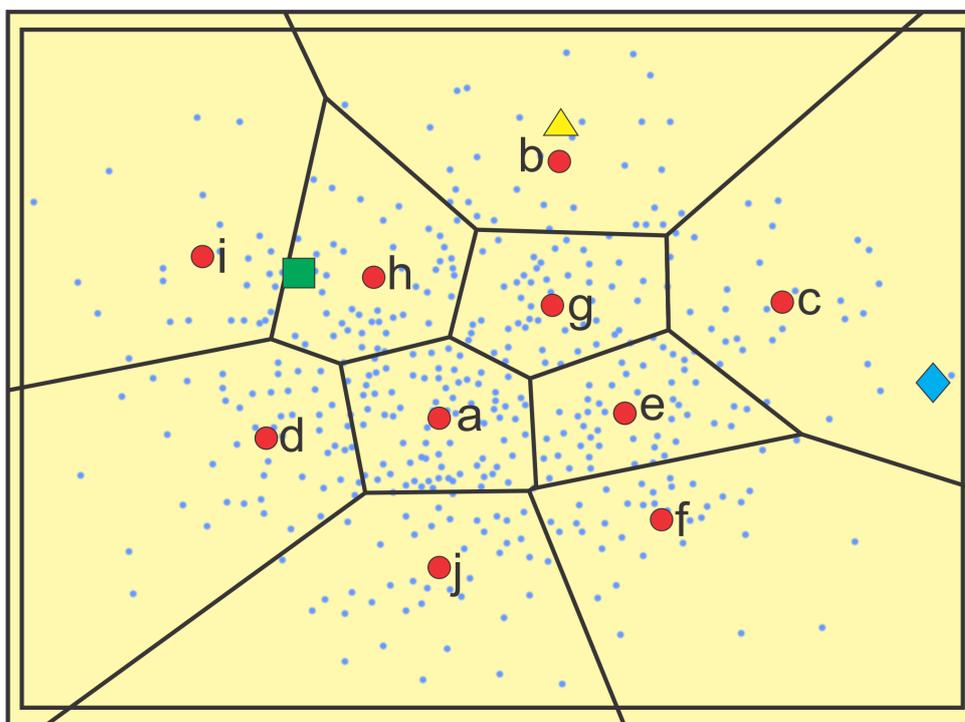


Figura 1.2: Os pontos pequenos representam descritores extraídos da imagem e os círculos maiores são centroides. O triângulo representa um descritor, que é adequado a rotulação pois está bem próximo a um centroide. O quadrado representa um descritor ambíguo pois se encontra em uma região de fronteira e o losango pode representar um ruído.

Com o objetivo de contribuir para a melhoria da produção agrícola de soja brasileira, este trabalho teve dois objetivos principais: teórico e prático. Do ponto de vista teórico, este trabalho avalia os problemas encontrados na construção dos descritores, permitindo que estes fossem mais representativos e efetivos. Para tanto, foram estudadas abordagens que incluem informações locais aos descritores, tais como as propostas por *Gemert et al* [63] que utiliza a atribuição leve ao vocabulário juntamente com a ambiguidade para se construir descritores que sejam mais discriminativos; e a proposta de *Jégou et al* [32] que utiliza a agregação de descritores de acordo com a distância às palavras visuais.

Do ponto de vista prático, este trabalho apresenta o desenvolvimento de um software para o reconhecimento de um conjunto de doenças foliares da cultura de soja, utilizando técnicas de visão computacional baseadas em descritores de pontos e histograma de palavras visuais. Essas informações tem como principal objetivo auxiliar agrônomos e produtores quanto a tomada de decisões visando o combate e a prevenção de doenças, tal como a aplicação de defensivo agrícolas em regiões precisas, a quantificação de danos e perdas para próximas safras, utilização de sementes tratadas, etc.

Este trabalho descreva uma abordagem que permite a atribuição de um descritor a mais do que uma palavra visual, através do cálculo de peso que o descritor exerce sobre a palavra

visual. O cálculo do peso leva em consideração a distância envolvida entre o descritor e as palavras visuais. Esse peso foi utilizado para a elaboração de um conjunto de técnicas que tomam como base o VLAD, são elas o VLAD-F que realiza a atribuição de cada descritor para todos os centroides, VLAD-PNC que atribui cada descritor a um conjunto P de centroides e VLAD-VNNC que atribui cada descritor a um conjunto variável de centroides. Também são descritos neste trabalho, um conjunto de experimentos e seus respectivos resultados, que tomou como base dois bancos de imagens compostos amostras de folhas de soja doentes e saudáveis, sendo que a primeira foi construída com o auxílio de um *scanner* e a segunda foi obtida através de uma câmera.

Nos experimentos foram utilizados dois dos descritores de pontos locais mais conhecidos na literatura, o SIFT e o SIFT Denso. Também foi utilizado ao algoritmo de agrupamento *K-means*, e para classificar as imagens foi utilizado o classificador *Linear Support Vector Machine* - SVM Linear [39], por ser simples, intuitivo e fornecer resultados fáceis de serem interpretados. Como métricas de avaliação foram calculadas as média dos percentuais de acerto, o desvio padrão e a matriz de confusão para as técnicas propostas.

De maneira geral as técnicas propostas obtiveram resultados superiores as existentes na literatura, para ambas as bases de imagens. E revelaram que a inclusão de informações locais aos descritores de características melhora o desempenho no reconhecimento de doenças foliares da cultura de soja. Além disso, especificadamente com relação a segunda base de imagens as técnicas demonstraram resultados promissores quanto a sua aplicação em problemas reais, visto que esta base representa de maneira mais fiel as condições existentes nos cultivos.

1.2 Trabalhos Correlatos

Nesta seção são descritos os principais trabalhos que estenderam o BoVW, construindo métodos mais representativos (veja Tabela 1.1).

Tabela 1.1: Trabalhos que promoveram extensões nas técnicas de classificação de imagens a partir do BoVW.

Autor	Ano	Título
Lazebnik, Svetlana and Schmid, Cordelia and Ponce, Jean	2006	Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories
Perronnin, Florent and Dance, Christopher and Csurka, Gabriela and Bressan, Marco	2006	Adapted vocabularies for generic visual categorization
Moosmann, Frank and Triggs, Bill and Jurie, Frederic	2007	Fast discriminative visual codebooks using randomized clustering forests
van Gemert, Jan C and Veenman, Cor J and Smeulders, Arnold WM and Geusebroek, J-M	2010	Visual word ambiguity
Jégou, Herve and Douzé, Matthijs and Schmid, Cordelia and Pérez, Patrick	2010	Aggregating local descriptors into a compact image representation
Wang, Jinjun and Yang, Jianchao and Yu, Kai and Lv, Fengjun and Huang, Thomas and Gong, Yihong	2010	Locality-constrained linear coding for image classification
Afonso, Luis Claudio Sugi	2012	Composição de dicionários visuais utilizando agrupamento de dados por Florestas de Caminhos Ótimos

Lazebnik *et al* [36] apresentam um método para realizar a categorização de imagens com base na correspondência geométrica global. Esta técnica funciona dividindo a imagem em sub-regiões cada vez menores e computando um histograma das informações locais encontradas dentro de cada sub-região. A pirâmide espacial resultante é uma extensão simples e computacionalmente eficiente do BoVW [15], que mostrou ser significativamente melhor na tarefa de categorização, fornecendo uma informação espacial grosseira ao BoVW. Em seus experimentos o método proposto se mostrou superior ao estado da arte, quando avaliado no banco de dados do Caltech-101 e atingiu alta precisão em um banco de quinze categorias de cenas [21]. Apesar da melhoria nos resultados, esta abordagem não é invariante a rotação, pois permite que um mesmo objeto esteja presente em uma imagem com orientações diferentes.

Perronnin *et al* [47] propuseram a realização de uma categorização visual genérica, *Generic Visual Categorization* (GVC), que é baseado na construção de um vocabulário universal que descreva todas as classes de imagens e de um conjunto de vocabulários das classes das imagens, obtido através da adaptação do vocabulário universal para uma classe específica de imagens. Onde cada histograma obtido, descreve se a imagem é apropriadamente modelada pelo vocabulário universal ou corresponde a um vocabulário adaptado específico. Perronnin *et al* [47] realizou experimentos em 3 bancos de dados diferentes, o banco de dados *in-house* de cenas, o banco de dados LAVA7 também conhecido como *Xerox7* [69] e a base de dados de *Wang* [11]. Os resultados obtidos demonstraram que essa abordagem de vocabulário universal ou vocabulário das classes correspondentes, superou as abordagens do estado da arte que caracterizam uma imagem por um único histograma.

Em seu trabalho, Moosmann *et al* [42] propuseram uma nova abordagem para a construção do vocabulário de palavras visuais através de uma floresta de agrupamentos randômicos. Esta abordagem se mostrou superior a codificação utilizando o *k-means*, pois forneceu um conjunto de palavras visuais mais discriminativo e resistente a ruídos. O método proposto primeiramente constrói um conjunto de árvores de decisão randômicas para predizer rótulos de classes para os vetores de descritores visuais e com os rótulos obtidos é construída a floresta. No método as árvores transformam cada descritor em um conjunto de índices de nós folhas de todas as árvores. Os votos para cada índice são acumulados para construir um histograma global que é utilizado na classificação.

Gemert *et al* [63] utilizam a chamada atribuição leve ao vocabulário visual que tem apresentado resultados superiores à atribuição rígida. Além disso, seu trabalho introduz o emprego da ambiguidade que também apresenta uma melhora no desempenho da identificação. Por meio da ambiguidade, tem-se um maior nível de representação de cada imagem, enriquecendo o vetor de características e acarretando em um melhor desempenho na etapa de classificação [63]. Ambiguidade pode ser modelada por diferentes métricas de similaridade, em que considera-se que um descritor pode ser semelhante a mais de uma palavra visual e tal semelhança pode ser representada por um intervalo normalizado entre -1 e 1. Em seu trabalho ele realizou a comparação entre quatro formas de atribuição: tradicional, aplicação de um Kernel, a ambiguidade *Codeword Uncertainty* e a ambiguidade *Codeword Plausibility*. A ambiguidade pela *Codeword Uncertainty* permite que um ponto seja atribuído a mais do que uma palavra visual com base no cálculo da silhueta. Já a ambiguidade pela *Codeword Plausibility* atribuem o melhor candidato com base na silhueta.

Jégou *et al* [32] propuseram o *Vector of Locally Aggregated Descriptors* (VLAD) que deriva do BoVW [15] e do kernel de *Fisher* [46], para fornece uma representação vetorial da imagem agregando os descritores com base em critérios de localidade. No método proposto os descritores são construídos a partir dos pontos de interesse da imagem, e um vocabulário de palavras visuais é obtido. Em seguida, cada descritor é associado a palavra visual mais pró-

xima através do método de vizinhos mais próximos, e é calculada a diferença entre o descritor e a palavra visual. A diferença entre cada palavra visual e seus descritores é acumulada para construir o descritor final. Em seus experimentos Jégou *et al* [32] demonstraram que o VLAD supera o BoVW para conjuntos de mesmo tamanho.

No trabalho de Wang *et al* [65], eles apresentam um esquema de codificação de características chamado *Locality-constrained Linear Coding* (LLC) que utiliza limitações de localidade para projetar cada descritor em seu sistema de coordenada local e é baseado no estudo feito por Yu *et al* [68]. Esta abordagem visa fornecer representações que funcionam melhores com classificadores lineares e permitem uma maior escalabilidade. Entre as principais propriedades do LLC estão a melhor reconstrução, dispersão local suave e solução analítica. Os resultados experimentais demonstraram que a representação final gerada por meio de código LLC conseguiu uma precisão de classificação de imagem notável, mesmo com um classificador SVM linear, além de ter apresentado baixa complexidade analítica.

Luis C. S. [2] propôs um método para encontrar o dicionário de maneira automática empregando um algoritmo baseado em grafos denominado Floresta de Caminhos Ótimos, o qual não necessita da dimensão do dicionário para encontrá-lo. Em seus experimentos Afonso utilizou 3 bases (*ETH80, Fish cells e Tropical Fruits*) e comparou as técnicas *k-means*, seleção aleatória e OPF não-supervisionado. Primeiramente foram extraídos os descritores SIFT [40] e SURF [5]. Após a extração dos descritores foram calculados os dicionários visuais utilizando as 3 técnicas e a avaliação foi realizada por meio de medições do tempo necessário para a geração de cada dicionário. Os resultados demonstraram que o algoritmo Floresta de Caminhos Ótimos é uma alternativa a ser empregada no BoVW, uma vez que as taxas de acerto são similares às demais técnicas, possui vantagem quando dicionários de alta dimensão devem ser calculados e, principalmente, não necessita que a dimensão do dicionário visual seja definido *a priori*.

1.3 Objetivos

1.3.1 Geral

Promover um conjunto de melhorias nos algoritmos de classificação de imagens que utilizam técnicas baseadas em descritores de pontos e histograma de palavras de visuais, possibilitando que estes sejam mais discriminativos e permitir a sua aplicação na identificação de doenças foliares na cultura da soja, como a Ferrugem Asiática, o Míldio, a Mancha Parda e a Mancha Olho de Rã.

1.3.2 Específicos

- Levantamento e construção de um banco de imagens com doenças foliares na cultura da soja;
- Propor um método que permita a inclusão de informações locais dos descritores de pontos através da atribuição leve;
- Implementação e validação do método proposto tendo como base o VLAD [32];
- Aplicação do método proposto no reconhecimento de doenças foliares na cultura de soja;
- Comparação dos resultados obtidos para o reconhecimento de doenças foliares da cultura da soja.

1.4 Estrutura do Texto

No Capítulo 1 foram apresentadas as motivações e objetivos deste trabalho. No Capítulo 2 é apresentada uma descrição sobre a soja, os ciclo de vida da planta, a descrição das doenças foliares da cultura de soja mais presentes no Brasil. O Capítulo 3 apresenta a Visão Computacional no contexto da agricultura e as principais técnicas envolvidas para realizar a extração de descritores locais, o SIFT [40] e o SIFT Denso [38]; o histograma de palavras visuais [15]; a técnica do VLAD.

No Capítulo 4 são descritas as abordagens propostas para solucionar o problema de inclusão de informações locais aos descritores, é apresentada uma solução baseada no cálculo de pesos proporcionais a distância envolvida entre o descritor e as palavras visuais, também são descritas as técnicas VLAD-F, VLAD-PNC e VLAD-VNNC que utilizam a atribuição de um descritor a mais do que uma palavra visual.

No Capítulo 5 são descritos os experimentos e resultados obtidos pelas técnicas propostas que utilizam a atribuição dos descritores com base no cálculo de pesos. Por fim, no Capítulo 6 são discutidas as conclusões deste trabalho e apresentados alguns trabalhos futuros.

A soja

A soja (*Glycine max L.*) é uma planta da família das *Fabaceae*¹ (leguminosas) originária da Ásia, e chegou ao Brasil por volta de 1882 [43] através dos imigrantes japoneses. Em 1914, ela foi oficialmente introduzida no Rio Grande do Sul, visto que o estado apresentava condições climáticas favoráveis ao desenvolvimento da cultura.

Porém, apenas na década de 60, a soja se estabeleceu como uma cultura economicamente importante para o Brasil, neste período cerca de 90% da produção brasileira se concentrava nos três estados da região Sul, em áreas onde prevalecia a combinação: trigo no inverno e soja no verão.

A expansão e o desenvolvimento efetivo da soja por todo o território nacional só ocorreu ao final década de 70, impulsionado pela indústria de óleo, pelas necessidades impostas pelo mercado mundial e também pela criação da Embrapa Soja em 1975, que tinha o propósito de desenvolver tecnologias para produção de soja no Brasil, tornando-se referência mundial em pesquisa para a cultura da soja em regiões tropicais [6].

O Centro-Oeste surgiu como uma nova opção produtiva da soja, a partir da década de 70, quando houve uma mecanização na agricultura. O cerrado, antes visto como um solo pobre, ganhou então um novo olhar. Dentre os fatores que contribuíram para o crescimento da soja no Centro-Oeste brasileiro [61], se pode destacar:

- incentivos fiscais disponibilizados para a abertura de novas áreas de produção agrícola, assim como para a aquisição de máquinas e construção de silos e armazéns;
- estabelecimento de agroindústrias na região, estimuladas pelos mesmos incentivos fiscais disponibilizados para a ampliação da fronteira agrícola;
- baixo valor da terra na região, comparado ao da Região Sul, nas décadas de 1960/70/80;
- desenvolvimento de um bem sucedido pacote tecnológico para a produção de soja na região, com destaque para as novas variedades adaptadas à condição de baixa latitude do Centro Oeste;

¹Representam uma grande e economicamente importante família das Angiospermas, que inclui as vagens, ervilhas e feijões

Embora, a produção de soja tenha crescido nos últimos anos, conforme Figura 1.1. O rendimento continua abaixo das expectativas, e entre os principais fatores que afetam negativamente a produção estão as doenças. Elas são responsáveis por grandes perdas na produção de diversos cultivos agrícolas, principalmente no que se referem a monoculturas maciças como a soja, cana-de-açúcar, arroz, laranja, entre outras.

Na natureza as espécies consideradas pragas encontram-se em equilíbrio com as demais populações, desta forma os danos causados não são graves a ponto de exigir qualquer método de controle, porém em cultivos de monoculturas isso não ocorre e existe um desequilíbrio que pode causar danos mais severos. Danos esses que vão desde a redução da produção, aumento de custo e do dano ambiental, pela necessidade de aplicação de defensivos agrícolas e outras medidas de manejo, bem como a baixa valorização dos produtos no mercado [25].

O sucesso no controle das doenças está intimamente ligado ao período da interferência realizada, seja ela por meio de medidas legislativas e tratos culturais que visam prevenir ou retardar a ocorrência de doenças para a safra seguinte ou por meio do controle por intervenções químicas aplicadas de forma preventiva ou curativa durante a safra [27].

No caso da soja, os períodos são representados pelos ciclos de vida da planta que são divididos em dois estágios: o vegetativo (V) apresentado pela Tabela 2.1 e que se inicia desde a sementeira até o florescimento [43], e o reprodutivo (R) ilustrado pela Tabela 2.2, que compreendem quatro fases distintas: o florescimento (R1 e R2) , desenvolvimento da vagem (R3 e R4), enchimento dos grãos (R5 e R6); e a maturação da planta (R7 e R8) [43]

A classificação dos estádios de desenvolvimento da soja é de suma importância para que ocorram intervenções de pesquisadores, extensionistas, agrônomos e produtores rurais de uma forma homogênea eliminando a possibilidade de erros [20].

Porém, definir quando uma medida deve ser tomada e em qual nível ainda é um problema, uma vez que o processo de detecção e reconhecimento de uma doença no meio agrícola é comumente realizado de forma visual por um especialista da área, que faz com que o processo seja lento e subjetivo. Por esse motivo, atualmente, agrônomos e agricultores optam por aplicar dosagens superiores visando não perder os investimentos feitos na plantação. Diante disso, estudos estão sendo feitos nos últimos anos visando auxiliar na tomada de decisões que permitam melhorar a produção e obter menor desperdício.

Tabela 2.1: Descrição dos estágios vegetativos da soja.

Estágio	Denominação	Descrição
VE	Emergência	Cotilédones acima da superfície do solo
VC	Cotilédone	Cotilédones completamente abertos
V1	Primeiro nó	Folhas unifoliadas completamente abertas
V2	Segundo nó	Primeira folha trifoliada completamente desenvolvida
V3	Terceiro nó	Segunda folha trifoliada completamente desenvolvida
V4	Quarto nó	Terceira folha trifoliada completamente desenvolvida
V5	Quinto nó	Quarta folha trifoliada completamente desenvolvida
V6	Sexto nó	Quinta folha trifoliada completamente desenvolvida
VN	Enésimo nó	Enésima folha trifoliada completamente desenvolvida

2.1 Doenças Foliares da Soja

A cultura da soja está sujeita a mais 40 doenças causadas por fungos, bactérias, nematoides e vírus. Esse número continua aumentando com a expansão da soja para novas áreas e, como consequência, do cultivo em monocultura. A importância econômica de cada doença

Tabela 2.2: Descrição dos estágios reprodutivos da soja.

Estágio	Denominação	Descrição
R1	Início do florescimento	Uma flor aberta em qualquer nó da haste principal
R2	Florescimento pleno	Uma flor aberta num dos dois últimos nós da haste principal, com folha completamente desenvolvida
R3	Início da formação de vagens	Vagem com 5 mm de comprimento num dos quatro últimos nós da haste principal, com folha completamente desenvolvida
R4	Vagem completamente desenvolvida	Vagem com 2 cm de comprimento num dos quatro últimos nós da haste principal, com folha completamente desenvolvida
R5	Início do enchimento de grãos	Grão com 3 mm de comprimento em vagem num dos quatro últimos nós da haste principal, com folha completamente desenvolvida
R6	Grão verde ou vagem cheia	Uma vagem contendo grãos verdes preenchendo as cavidades da vagem de um dos quatro últimos nós da haste principal, com folha completamente desenvolvida
R7	Início da maturidade	Uma vagem normal na haste principal com coloração de madura
R8	Desfolha natural	Plantas com vagens amarelas e 50% de desfolha
R9	Maturidade plena	95% das vagens com coloração de madura

varia de ano para ano e de região para região, dependendo das condições climáticas de cada safra e das medidas de manejo de doenças empregas ao longo dos anos.

A relação para a ocorrência de doenças em ambientes agrícolas pode ser explicado pela Figura 2.1. Hospedeiro, patógeno e ambiente são representados por cada lado do triângulo, onde a ocorrência de doença depende da combinação simultânea desses três fatores. Como por exemplo, a falta de condições meteorológicas para o surgimento da ferrugem asiática. O patógeno pode estar presente na área de cultivo, bem como os hospedeiros, mas sem as condições ambientais adequadas, a interação da planta com o patógeno não resulta em doença. Nesta seção, serão apresentadas as principais doenças foliares que ocorrem na cultura da soja na região Centro-Oeste do Brasil [25].

2.1.1 Ferrugem Asiática (*Phakopsora Pachyrhizi*)

A ferrugem asiática da soja é uma das doenças mais severas que ocorre na cultura da soja, com danos variando de 10 a 90% nas diversas regiões geográficas em que ocorre, segundo Yorinori *et al.* [27]. Os sintomas iniciais da doença são pequenas lesões foliares, de coloração castanha a marrom escura. Na face inferior da folha, pode-se observar urédias² que se rompem e liberam os uredósporos, conforme Figura 2.2. Plantas severamente infectadas apresentam desfolha precoce, que compromete a formação, o enchimento das vagens e o peso final do grão. Quanto mais cedo ocorrer a desfolha, menor será o tamanho do grão e, consequentemente, maior a perda de rendimento e de qualidade [66].

A infecção ocorre sob temperaturas entre 15°C e 28°C e elevadas umidades relativas (75 a 80%), segundo Juliatti *et al* [33]. Ambientes com períodos prolongados de orvalho e umidade

²Estruturas de reprodução do fungo

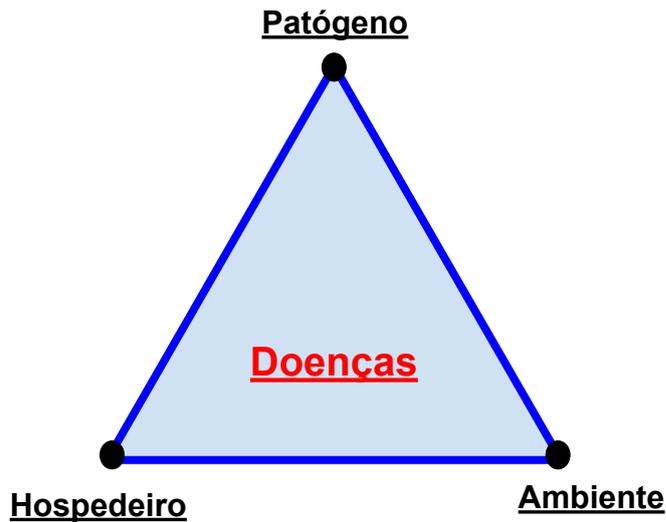


Figura 2.1: Diagrama esquemático das inter-relações dos fatores envolvidos em epidemias de doenças de plantas. Adaptado de [25].

são favoráveis para o progresso da doença na lavoura. Diferente de outras doenças, a ferrugem asiática da soja não penetra pelos estômatos ou por ferimentos, ela penetra diretamente através da cutícula e epiderme, tornando a infecção mais rápida e fácil [25].



(A)



(B)

Figura 2.2: Sintomas da ferrugem asiática em folhas de soja. (A) ferrugem RB e em (B) ferrugem TAN. As lesões RB são caracterizadas por uma coloração marrom avermelhadas e são consideradas um tipo de lesão mais resistente quando comparadas com as lesões TAN.

O controle da ferrugem asiática da soja exige a combinação de diversas técnicas, a fim de evitar perdas de rendimento. A Fundação MS [25], recomenda algumas estratégias, tais como: semear preferencialmente cultivares precoces e no início da época recomendada para cada região; evitar o prolongamento do período de semeadura, pois a soja semeada mais tardiamente (ou de ciclo longo) irá sofrer mais dano, devido à multiplicação do fungo nas primeiras semeaduras. Nas regiões onde foi constatada a ferrugem, deve-se iniciar a vistoria da lavoura desde o início da safra e, principalmente, quando a soja estiver próxima da floração, ao primeiro sinal da doença e, havendo condições favoráveis (chuva e/ou abundante formação

de orvalho), poderá haver a necessidade de aplicação de fungicida. O monitoramento contínuo é essencial para que a medida de controle possa ser adotada no momento correto, a fim de evitar reduções de produtividade.

2.1.2 Mancha Alvo (*Corynespora Cassiicola*)

No Brasil, o desenvolvimento da mancha alvo nos campos de soja ganhou destaque nos últimos anos. Além disso, esta doença já foi relatada em algumas espécies de plantas daninhas [58]. O fungo é encontrado em praticamente todas as regiões de cultivo de soja do Brasil, acreditando-se ser nativo e infectar um grande número de espécies de plantas. Pode sobreviver em restos de cultura e sementes infectadas, sendo essa uma forma de disseminação. Segundo a Fundação MS [25], condições com alta umidade relativa e temperaturas amenas são favoráveis à infecção na folha.

Os sintomas mais comuns são manchas nas folhas, com halo amarelado e pontuação escura no centro, que causam severa desfolha, conforme Figura 2.3. Podendo ocorrer também manchas na haste e na vagem. O fungo pode infectar raízes, causando podridão radicular e intensa esporulação [27].

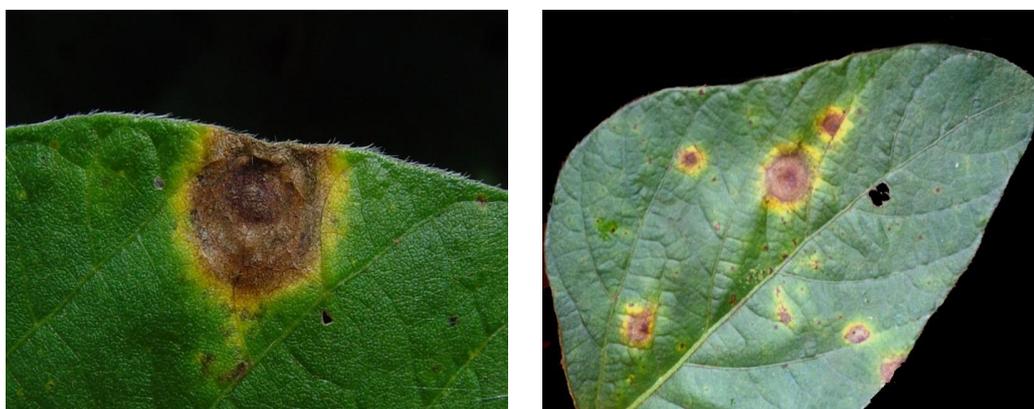


Figura 2.3: Sintomas da Mancha alvo em folhas de soja. Fonte <http://dirceugassen.com/>.

Várias estratégias são recomendadas para o controle da doença tais como: o uso de cultivares resistentes, o tratamento de sementes, a rotação/sucessão de culturas com milho e espécies de gramíneas e pulverizações com fungicidas [27].

2.1.3 Míldio (*Peronospora Manshurica*)

O míldio da soja é uma doença amplamente disseminada no Brasil [57], e estima-se que os prejuízos decorrentes desta doença estão entre 8 e 14% quando ocorre em cultivares suscetíveis [18]. As condições que favorecem a ocorrência desta doença são elevados períodos de molhamento foliar (12 horas) e temperaturas entre 20 °C e 22 °C durante qualquer estágio fenológico da cultura [50].

Os sintomas da doença são bem característicos, com manchas de coloração verde-claras até amareladas, localizadas na face adaxial das folhas. Com o desenvolvimento da doença, as lesões se tornam marrom-acinzentadas até marrom escura, com margens amareladas ou esverdeadas. O crescimento do fungo pode ser observado em algumas situações, de coloração cinza na face abaxial da folha, conforme Figura 2.4, principalmente em períodos úmidos.

Folhas severamente atacadas se tornam marrom e podem se soltar da planta, causando desfolha. Também pode se observar o desenvolvimento esbranquiçado do fungo no interior

de vagens e na superfície das sementes. As plântulas originadas de sementes infectadas são raquíticas com folhas manchadas.



Figura 2.4: Sintomas de mildio em folhas de soja.

As sementes contaminadas são a principal forma de disseminação do patógeno e fonte de inóculo inicial. As sementes infectadas com o patógeno apresentam redução do peso e do teor de proteína, além de atraso na germinação das plântulas. Segundo a Fundação MS [25], o combate ao mildio envolve a rotação de culturas com espécies não hospedeiras do patógeno, destruição dos restos culturais, uso de cultivares resistentes, e o tratamento de sementes com fungicidas, sendo a última a mais eficiente das estratégias.

2.1.4 Mancha Olho de Rã (*Cercospora Sojina*)

A mancha-olho-de-rã afeta todas as partes aéreas da planta, comprometendo os rendimentos e a qualidade dos grãos. O fungo produz sintomas nas folhas, hastes, vagens e sementes. A doença ocorre em todas as regiões do mundo produtoras de soja, com grande incidência sobre *Glycine max* no Brasil, Camarões, Canadá, China, Cuba, Índia, Japão, México, Nepal, Taiwan, USA e Zâmbia.

Entre os danos, podem ocorrerem folhas, hastes, vagens e sementes. As lesões iniciam como pequenos pontos de encharcamento e evoluem para manchas castanho-claras no centro com bordos castanho-avermelhados, na página superior da folha, e cinzas na página inferior, onde ocorre a esporulação. O tamanho das lesões varia de 1 a 5 mm de diâmetro. Em hastes e vagens, as lesões aparecem ao final da granação e apresentam aspectos de anasarca na fase inicial, evoluindo para manchas circulares, castanho-escuras nas vagens, e manchas elípticas ou alongadas, com centro cinza a castanho-claro e bordos castanho-avermelhados na haste. Na semente, o tegumento apresenta rachaduras e manchas de tamanhos variáveis, de coloração parda a cinza. De modo geral, a ocorrência da mancha olho-de-rã começa na fase de floração, entre os estádios R1 e R3.

O controle mais eficiente e econômico é o uso de variedades resistentes. O controle químico pode ser realizado com apenas duas aplicações de fungicidas: a primeira quando houver 5 a 10 manchas nas folhas mais afetadas e a segunda, se necessária, entre 10 e 15 dias após.

2.1.5 Mancha Parda (*Septoria Glycines*)

É uma doença amplamente disseminada no país, podendo causar grandes perdas no rendimento de grãos. A doença é inicialmente observada nas folhas primárias, causando lesões circulares, de cor vinho, que posteriormente tornam-se necróticas. Para a soja, é uma das



Figura 2.5: Sintomas da mancha-olho-de-rã em folhas de soja. Fonte <https://www.agrolink.com.br/>.

primeiras doenças a aparecer nos campos de cultivo, manifestando-se com maior intensidade a partir do estágio R5 (início de formação de vagens), associando-se mais tarde à infecção por *Cercospora kikuchii* [48]. Em períodos de seca, pode haver aumento de desfolha, especialmente de folhas localizadas no terço inferior das plantas.



Figura 2.6: Sintomas da Mancha Parda em folhas de soja. Fonte <https://www.agrolink.com.br/>.

O controle ideal da mancha parda seria através de genótipos resistentes. Infelizmente, para a cultura da soja, nenhum genótipo com essa característica foi encontrado, sendo observadas apenas diferenças em suscetibilidade entre genótipos.

Referencial Teórico

Visão computacional é a ciência responsável pela visão e pela forma como um computador enxerga o meio físico a sua volta, extraindo informações significativas a partir de imagens obtidas com câmeras, sensores, satélites e outros dispositivos [4]. Com base nessas informações e em modelos representativos, o computador adquire a capacidade de reconhecer, manipular e interagir de diversas formas com o meio físico que compõe as imagens. Uma das principais aplicações da visão computacional é a classificação de imagens, com base em informações, atributos e dados obtidos das imagens.

A classificação de imagens é fundamental em diversos problemas de Visão Computacional, como reconhecimento de objetos e de cenas [56], montagem automática de mosaicos [64], geração de objetos 3D a partir de várias imagens [13], análise de movimentos [51], etc. Uma abordagem para classificar imagens é a utilização de descritores para identificar características de uma imagem e reunir informações sobre as mesmas. A forma como uma imagem é descrita depende da aplicação, e pode ser com descritores baseados em cor, pontos de interesse, detectores de bordas, histogramas, entre outros. O SIFT (*Scale Invariant Feature Transform* [40]) é um descritor de pontos de interesse que juntamente com o histograma de palavras visuais (*Bag-of-Visual-Words*) busca pontos nas imagens a fim de encontrar e descrever regiões que tenham maior relevância.

A Seção 3.1 descreve a visão computacional e seus principais conceitos. Na Seção 3.2 descreve o descritor de pontos locais SIFT que utiliza informações extraídas a partir de um espaço de escalas. Na Seção 3.3 é descrito o SIFT Denso, uma variante do SIFT que elimina a etapa de detecção de pontos e utiliza uma grade densa de pontos de interesse. Na Seção 3.4 é apresentado o BoVW que utiliza os pontos descritos pelo SIFT para a construção de um histograma que é utilizado na classificação das imagens. A Seção 3.5 apresenta o VLAD (*Vector of Locally Aggregated Descriptors*) [32], que utiliza vetores para agregar os descritores de uma imagem como base em critérios de localidade.

3.1 Visão computacional e sua aplicação na agricultura

Segundo Maia [41], a visão computacional é o ramo da Ciência da Computação que reúne as teorias e tecnologias desenvolvidas com a finalidade de que imagens sejam interpretadas

por sistemas artificiais implementados em computadores. Segundo Prince [54], o principal objetivo da visão computacional é extrair informações úteis das imagens. No entanto, esta tarefa tem se mostrado profundamente desafiadora e despertado o interesse de diversos pesquisadores nas últimas décadas.

A visão computacional é uma área de pesquisa que pode incluir métodos de aquisição de imagens, pré-processamento, segmentação, extração de atributos ou características e reconhecimento de padrões. Por exemplo, técnicas de processamento e análise de imagens possibilitam extrair informações visuais como, forma, cor e textura, com o objetivo de identificar características das imagens para a utilização em um sistema de reconhecimento de padrões [7].

Em geral, a visão computacional pode resolver problemas de qualquer área de pesquisa que envolva imagem. Atualmente, é utilizada em uma grande quantidade de aplicações, tais como: reconhecimento de caracteres e de objetos, sensores de movimento, vigilância, biometria, controle de qualidade, robótica, medicina e agricultura.

Na agricultura, devido a busca por maior competitividade, as exigências do mercado consumidor e a necessidade de preservação dos recursos naturais tem impulsionado o setor a buscar maneiras de aumentar a produtividade e ainda diminuir o impacto ambiental. Vários projetos que utilizam técnicas de Visão Computacional direcionados à agricultura vêm sendo implementados para a obtenção de melhores resultados, aumento da produção, qualidade dos produtos, redução de desperdícios, entre outros.

Dentre os trabalhos desenvolvidos na Visão Computacional voltado para a agricultura, destacam-se aqueles utilizados na agricultura de precisão para a tomada de decisões e automatização de tarefas através do processamento de informações obtidas em imagens, como o sensoriamento remoto para o mapeamento de áreas de cultura e estimação do rendimento, análise do comportamento animal, análise de imagens fornecidas por VANTs ¹. Outro campo que ganhou destaque nos últimos anos é o monitoramento das condições da cultura como, por exemplo, a identificação das deficiências nutricionais das plantas, e mais recentemente a ocorrência de doenças, classificação de plantas invasoras, dentre outros [23].

Em Nazaré Júnior *et al* [17] foi proposto o desenvolvimento de um sistema de detecção de área foliar danificada em folhas de soja utilizando segmentação de imagens pelo algoritmo de limiarização automática de Otsu [62], que propõe a construção de um histograma da imagem com duas classes: objeto e fundo. Em outro trabalho, Souza [59] propôs a construção de um sistema automatizado para a identificação de leveduras viáveis na cana-de-açúcar que são importantes para a produção de etanol, através do histograma de palavras visuais e de aprendizado supervisionado.

Em seu trabalho Santos *et al* [55] propôs o desenvolvimento de um sistema para o reconhecimento de plantas invasoras e percentual de ocupação utilizando técnicas de processamento digital de imagens e reconhecimento de padrões. Esse sistema utiliza imagens georreferenciadas, adquiridas por uma câmera digital que são pré-processadas. Em seguida, utilizando-se técnicas de segmentação baseadas em texturas são selecionadas as plantas invasoras, a partir de então são construídos descritores baseados nas formas das plantas invasoras. Por fim, através do reconhecimento de padrões são obtidos dois mapas, um composto por folhas largas e outro com folhas estreitas.

Cruvinel *et al* [14] propôs a construção de mapas baseados em técnicas de visão computacional para a aplicação de herbicidas em taxas variáveis dedicados a plantas invasoras de folha longa e estreita na cultura de milho. O referido modelo considera a etapa de coleta de informações georreferenciadas, pré-processamento para quantificar informações de interesse,

¹Veículos Aéreos Não Tripulados

construção de mapas de textura, extração de características geométricas e a geração de mapas de plantas invasoras de folhas largas e estreita. Com base nos mapas é possível realizar a aplicação de herbicidas de forma localizada e em taxas variáveis que podem minimizar os impactos econômicos e ambientais.

Em geral, as soluções desenvolvidas na visão computacional são baseadas em técnicas e modelos de processamento de imagens, que permitem a extração e descrição de informações a partir de imagens. A forma como as informações são obtidas e descritas são um passo fundamental dos processos, e podem ser a partir de pontos de interesse, bordas, cores, texturas, entre outros. A seguir são apresentados dois dos principais descritores de pontos estudados na literatura, SIFT [40] e o SIFT Denso [38].

3.2 SIFT

O SIFT (*Scale Invariant Feature Transform*)[40] é uma técnica de processamento de imagens proposta por David Lowe. Tem sua origem na necessidade de lidar com o reconhecimento e classificação de objetos em imagens, onde podem haver interferências, como: mudança de escala, alteração na iluminação, mudança no ângulo da imagem, etc. O SIFT se divide em duas partes: Detector e Descritor.

- **Detector:** É baseado no cálculo da diferença de Gaussianas para localizar pontos de interesse;
- **Descritor:** Utiliza histogramas de gradiente orientados a fim de descrever uma vizinhança local dos pontos de interesse.

O SIFT extrai um conjunto de informações para descrever localmente cada ponto de interesse, que tem como principais características: alta distintividade; invariância a escala, rotação, parcialmente a iluminação e variações 3D. Esse conjunto de informações, dá-se o nome de descritor local. Essa extração através do SIFT é realizada em 4 passos, sendo os dois primeiros correspondentes ao detector e os dois últimos ao descritor:

1. Detecção de extremos;
2. Localização e eliminação de pontos;
3. Definição da orientação dos pontos;
4. Construção do descritor.

3.2.1 Detecção de extremos

A primeira etapa do SIFT consiste na detecção de pontos que sejam invariantes a mudanças de escala da imagem. Tal objetivo é alcançado procurando por características estáveis em diferentes escalas, por meio da construção de um espaço de escalas, que neste caso é realizada através da função Gaussiana [37].

Uma imagem $I(x,y)$ passa a ser representada por $L(x,y,\sigma)$ no espaço de escala. Esta representação é obtida ao aplicar um núcleo de convolução Gaussiano $G(x,y,\sigma)$ sobre a imagem $I(x,y)$ conforme:

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y) \quad (3.1)$$

em que o núcleo Gaussiano $G(x,y,\sigma)$ é definido segundo a Equação 3.2:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \quad (3.2)$$

É possível observar que este núcleo representa uma escala através do desvio padrão σ . A convolução de um núcleo Gaussiano aplicado em cada imagem faz com que a imagem fique suavizada. Com isso, são eliminados detalhes da imagem, incluindo ruídos, restando apenas as características globais. Ao variar a escala (σ) as características globais relacionadas com a escala σ são evidenciadas. Logo, utilizando apenas a função Gaussiana, variando a escala, é obtida uma quantidade enorme de pontos de extremo com baixa distintividade. Portanto, para obter uma maior eficiência na localização de pontos de interesse no espaço de escala, faz-se o uso da função de Diferenças de Gaussianas (*Difference of Gaussian* - DoG) [24].

A função DoG é composta pela diferença de imagens filtradas em escalas próximas, separadas por uma constante de escala k . A função DoG é definida pela Equação 3.3.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) \otimes I(x, y) \\ D(x, y, \sigma) &= G(x, y, k\sigma) \otimes I(x, y) - G(x, y, \sigma) \otimes I(x, y) \\ D(x, y, \sigma) &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (3.3)$$

A imagem $D(x, y, \sigma)$ os pixels com valores altos correspondem a regiões onde há mudanças bruscas de intensidade.

Nas Figura 3.1(a) apresenta uma imagem em tons de cinza, enquanto que na Figura 3.1(b), 3.1(c) e 3.1(d) é possível observar a aplicação do núcleo Gaussiano em diferentes escalas σ e perceber o efeito de suavização conforme o aumento da escala, σ . Na Figura 3.1(e) e 3.1(f) é apresentada a diferença entre Gaussianas, levando-se em consideração os resultados da aplicação do filtro Gaussiano anterior.

A Figura 3.2 mostra a construção do espaço de escalas, este esquema é composto por quatro passos e tem como resultado a geração das oitavas. Em cada oitava obtêm-se os conjuntos de imagens L e D para a imagem analisada, para tornar o método invariante a escalas diferentes.

Para a obtenção do espaço de escalas, os seguintes passos são executados:

1. A imagem original sofre convoluções com filtros Gaussianos para a geração de imagens separadas por um fator de escala k no espaço de escala, conforme a Figura 3.2(a);
2. Lowe [40] considera necessário fazer convoluções da imagem até 2σ em cada oitava para ser possível a construção de descritores invariantes a escala. Portanto, para se gerar s imagens convoluídas, o fator de escala k é definido por $k = 2^{1/s}$;
3. Imagens adjacentes na Figura 3.2(a) são subtraídas para produzir as imagens da Diferença de Gaussianas, conforme mostrado na Figura 3.2(c);
4. O processo se repete reduzindo-se a resolução da imagem pela metade, conforme apresentado na Figura 3.2(b), em cada oitava a imagem é convoluída por filtros Gaussianos em diferentes escalas, criando um conjunto de $s + 3$ imagens convoluídas por oitava de maneira que a detecção de extremos cubra toda oitava.

A partir do espaço de escalas, é realizada a detecção de extremos em cada intervalo contendo todos os vizinhos do espaço de escala. Os extremos são dados por valores de máximo ou mínimo locais para cada $D(x, y, \sigma)$, que são obtidos comparando-se o valor de cada ponto com o valor de seus 8 vizinhos na sua escala, com os nove vizinhos na escala superior, e os nove vizinhos na escala inferior, conforme apresentado na Figura 3.3. Nesta figura, o ponto "X" é comparado com seus vizinhos em três imagens.

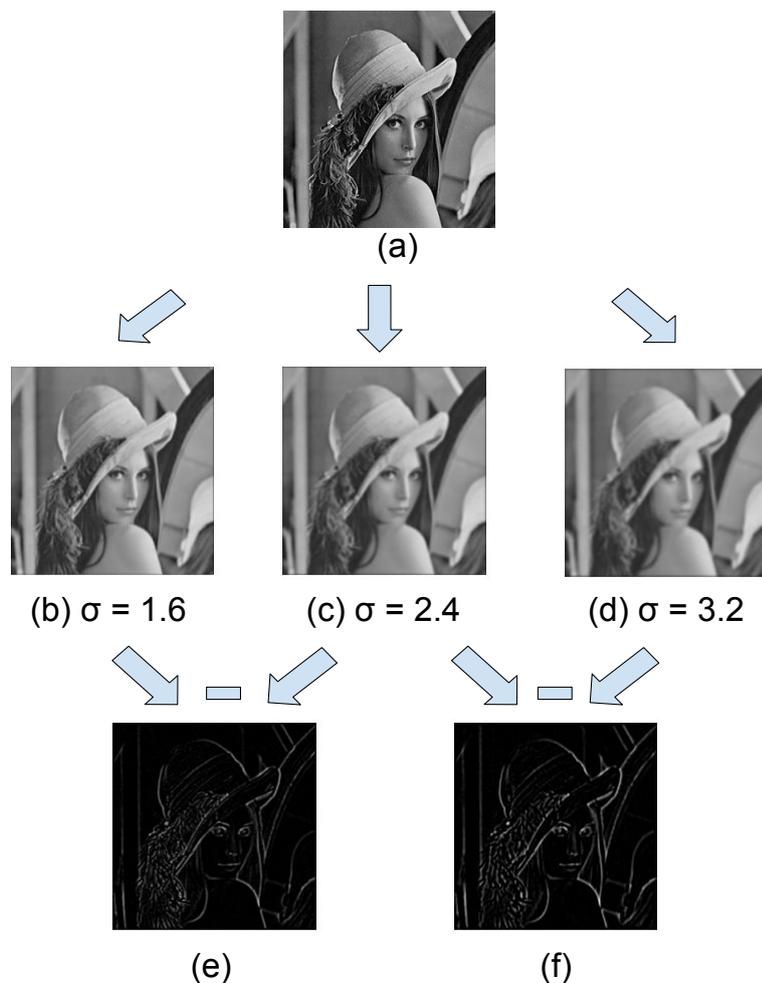


Figura 3.1: (a) Imagem original em escala de cinza. Em (b), (c) e (d) observa-se a convolução de (a) com núcleos Gaussianos com diferentes valores de σ . É possível observar que a medida que σ aumenta, a imagem fica mais suavizada. Finalmente, (e) e (f) apresentam as diferenças de Gaussianas.

Após a detecção de pontos extremos, máximos e mínimos, o SIFT retorna um conjunto de pontos p_i de potenciais pontos de interesse, sendo necessário determinar a localização exata e o descarte de pontos de interesse instáveis.

3.2.2 Localização e eliminação de pontos

Todos os pontos p_i que foram identificados como extremos são candidatos a ponto de interesse, sendo necessário determinar a localização precisa e eliminar pontos instáveis. O método consiste em ajustar uma função quadrática 3D do ponto de amostragem local de modo a determinar a localização interpolada do máximo.

Isto é realizado aplicando a expansão de Taylor da função DoG sobre a imagem $D(x, y, \sigma)$, deslocada de tal forma que a origem desta expansão seja o ponto de amostragem [40], conforme a Equação 3.4:

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \bar{x}} \hat{x} \quad (3.4)$$

Segundo Brown and Lowe [9], deve-se rejeitar valores de $|D(\hat{x})|$ inferiores a um determinado limiar. Ele considera apropriado trabalhar com um limiar igual 0.03, assumindo-se que os tons de cinza dos *pixels* da imagem estejam normalizados em valores entre 0 e 1.

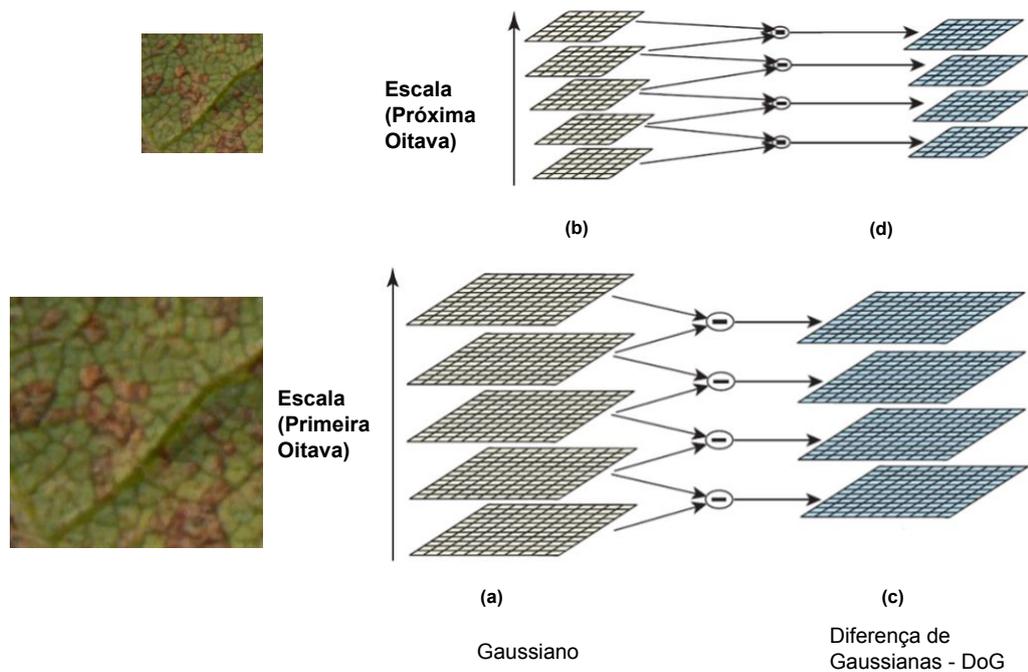


Figura 3.2: Representação do procedimento de obtenção das Diferenças de Gaussiano DoG para diversas escalas de uma imagem. Imagem baseada em [40] e adaptada utilizando imagens de doenças foliares da soja.

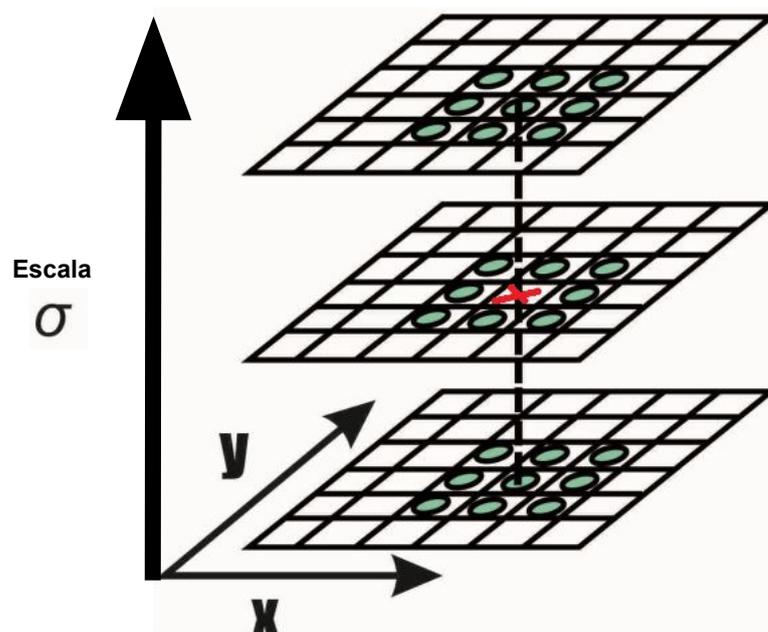


Figura 3.3: Detecção de extremos no espaço-escala σ . Imagem obtida e adaptada de [40].

Mesmo após a eliminação de pontos de interesse com baixo contraste, alguns pontos ainda precisam ser eliminados, visando estabilidade. A diferença de Gaussiano (DoG) resulta em valores altos para os pontos de interesse localizados em bordas, inclusive para aqueles em uma localização ruim e sensível a ruídos. Com base no detector de cantos de Harris (*Harris corner detector*) [26], é calculada a proporção entre as principais curvaturas. Isso é possível utilizando uma matriz Hessiana 2×2 , H , calculada na localização e escala do ponto de interesse, conforme Equação 3.5:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3.5)$$

onde D_{xy} é a derivada de $D(x, y, \sigma)$ na localização e escala em relação a x e y ; D_{xx} é a derivada segunda em relação a x ; e D_{yy} é a derivada segunda em relação a y .

A Hessiana representa assim uma segunda derivada, permitindo mensurar as magnitudes das curvaturas de D a partir de seus autovalores. Determina-se α , o autovalor com maior magnitude, e β , o de menor. Podendo-se então calcular a soma dos autovalores pelo traço de H e o produto pelo seu determinante:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (3.6)$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (3.7)$$

Para o caso em que o determinante for negativo, as curvaturas possuem sinais diferentes, e o ponto é descartado, não sendo considerado um extremo. Sendo r a razão entre o autovalor de maior magnitude e o de menor, de modo que $\alpha = r\beta$, então:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \quad (3.8)$$

A Equação 3.8 depende apenas da razão entre os autovalores, sendo independente de seus valores individuais. O valor de $(r+1)/r$ oferece uma medida de quanto os autovalores são distintos, ou seja, é mínimo quando são idênticos e cresce com respeito ao valor de r . Desta forma, são eliminados pontos indesejáveis abaixo de um determinado limiar(r), que estão próximos as extremidades, conforme:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \quad (3.9)$$

Segundo os resultados obtidos por Lowe [40], a resposta da principal curvatura será alta quando computada com *pixels* que atravessam uma borda e será baixa quando os *pixels* estiverem na direção da borda. Essa informação é relevante pois, as derivadas da matriz Hessiana H são computadas através da subtração de *pixels* próximos aos pontos de interesse.

A Figura 3.4, ilustra os passos da extração dos pontos de interesse e a remoção de pontos com baixo contraste e baixa resposta em bordas. Na Figura 3.4(a) temos os pontos extraídos com o DoG, em estado da arte, na Figura 3.4(b) são mostrados os pontos restantes após a eliminação de pontos com baixo contraste. Na Figura 3.4(c) são apresentados os pontos restantes após a eliminação de pontos com resposta baixa em bordas.

3.2.3 Definição da orientação dos pontos

Com o objetivo de obter a invariância à rotação, é definida uma orientação para cada ponto de interesse avaliando a região ao seu redor. Calcula-se para cada ponto de interesse de imagem na escala, $L(x, y, \sigma)$, a magnitude $m(x, y)$ e orientação $\theta(x, y)$ do gradiente usando as diferenças entre a intensidade dos *pixels*, conforme Equação 3.10 e Equação 3.11, respectivamente:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3.10)$$

$$\theta(x, y) = \tan^{-1} \left(\frac{(L(x, y+1) - L(x, y-1))}{(L(x+1, y) - L(x-1, y))} \right) \quad (3.11)$$

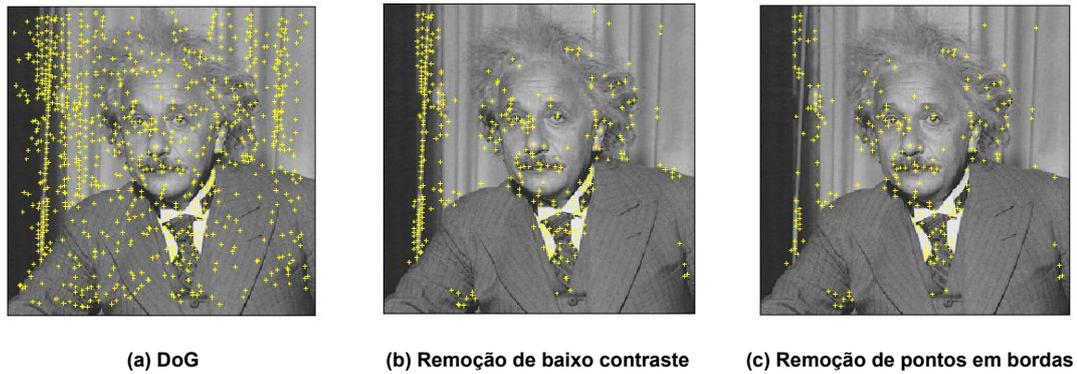


Figura 3.4: (a) Pontos obtidos a partir do DoG, (b) Pontos restantes após a eliminação de pontos com baixo contraste, (c) Pontos restantes após a eliminação de pontos em bordas. Imagem obtida e adaptada de [19].

Utilizando as orientações do gradiente ao redor do ponto de interesse, monta-se um histograma de orientações a cada 10 graus, totalizando 36 faixas, cobrindo todas as orientações possíveis (0 a 360 graus). Para cada pixel, é atribuído um peso definido pela sua magnitude do gradiente e pelo peso da função Gaussiana, conforme Figura 3.5(a). Cada magnitude e orientação calculada ao redor do ponto de interesse é adicionada ao histograma. A orientação determina a posição em que será adicionada no histograma e a magnitude, o valor a ser somado, conforme Figura 3.5.

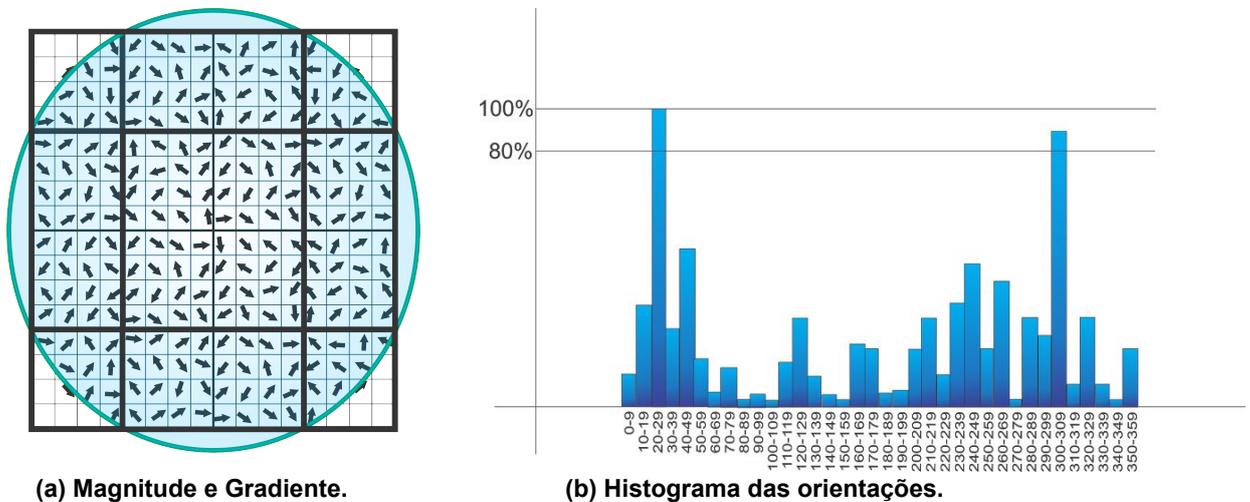


Figura 3.5: (a) Região delimitada por uma função de peso Gaussiano, com círculo. Imagem adaptada de [40]. (b) Representação de um possível histograma utilizado para determinar orientação de um ponto de interesse. Imagem adaptada de [60].

Os picos no histograma de orientações indicam direções dominantes dos gradientes locais. O maior pico do histograma determinará a orientação θ_i do ponto de interesse em questão, e qualquer outro pico cujo valor seja maior ou igual a 80% do valor do maior pico, será utilizado para criar outro ponto de interesse com a mesma posição e escala, porém com a orientação determinada por esse outro pico. Na Figura 3.6 é apresentado um conjunto de pontos de interesse extraídos e suas respectivas orientações.

Ao final dessas três primeiras etapas, cada ponto de interesse é identificado pelas seguintes informações:

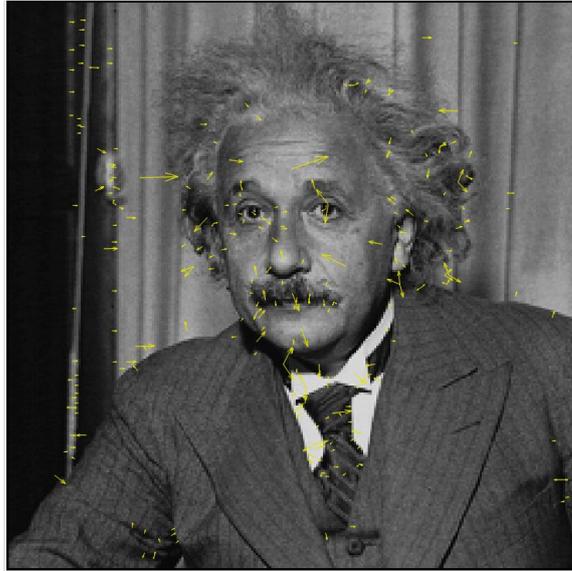


Figura 3.6: Setas em amarelo indicam a orientação do ponto de interesse. Imagem adaptada de [19].

- localização espacial (x_i, y_i) ;
- escala σ_i em que foi detectado;
- orientação θ_i .

3.2.4 Construção do descritor

As operações descritas anteriormente atribuíram a cada ponto de interesse quatro valores: as coordenadas (x_i, y_i) de sua posição; a escala σ_i ; e a orientação θ_i . O próximo passo é construir um descritor para a região local da imagem que seja invariante a mudanças de escala, iluminação e ponto de vista 3D, tornando-os distinguíveis.

Os procedimentos a seguir são executados com os valores normalizados em relação à orientação definida na seção anterior para cada ponto de interesse. Para que os descritores sejam invariantes à rotação, as orientações dos gradientes destes pontos são rotacionadas em um ângulo correspondente à orientação do ponto de interesse obtida na seção anterior.

O descritor do ponto de interesse é criado computando-se as magnitudes e orientações dos gradientes que são amostradas ao redor da localização do ponto de interesse. Este procedimento pode ser observado na Figura 3.7(a), onde os gradientes são representados pelas setas. A região ao redor do ponto de interesse é dividida em uma grade com $n \times n$.

Uma função Gaussiana é utilizada para dar peso à magnitude do gradiente em cada ponto na vizinhança do ponto de interesse, com uma janela de suavização Gaussiana de escala σ igual à metade da largura da janela do descritor. O propósito dessa janela Gaussiana é evitar mudanças súbitas do descritor fazendo pequenas mudanças na posição da janela, e também dar menos ênfase aos gradientes que estão longe do centro do descritor. Então é calculado um histograma para cada quadrante, com base nas magnitudes dos gradientes dos pixels de cada região.

O descritor é formado pela concatenação dos histogramas de cada grade, gerando um único descritor para cada ponto de interesse de 128 dimensões. Conforme ilustrado na Figura 3.7 (b), são 4×4 regiões por $\times 8$ orientações possíveis. Ao final, o descritor de características é normalizado para reduzir as influências geradas por mudança de iluminação.

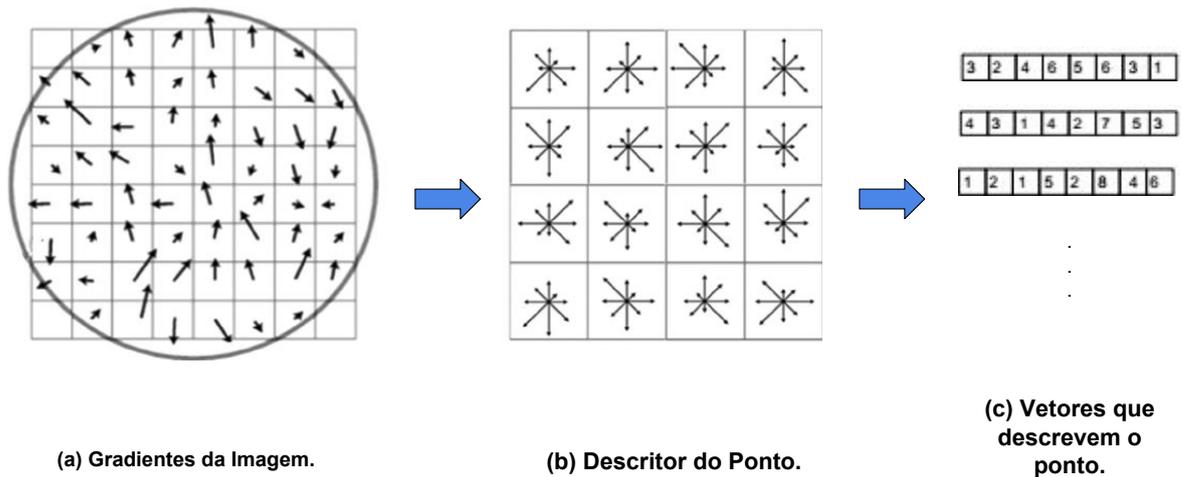


Figura 3.7: Representação da janela de gradientes e do descritor de ponto de interesse obtido através do SIFT. Imagem adaptada de [24].

Ao final das 4 etapas, o SIFT detecta M pontos de interesse com as seguintes propriedades: $\varphi_i = [x_i, y_i, \sigma_i, \theta_i, r_i]$, sendo (x_i, y_i) a posição espacial do ponto, σ_i é a escala, θ_i a orientação e r_i são os descritores.

3.3 SIFT Denso

Após os bons resultados do SIFT, diversas variantes foram propostas, como PCA-SIFT [34] e SIFT Denso [38]. O SIFT Denso é um descritor que tem apresentado bons resultados nas tarefas de reconhecimento de cenas e objetos [38], também demonstrou resultados expressivos no trabalho proposto por [28] para estimativa automática da pose da cabeça e também na tarefa de detecção de doenças da soja no estudo proposto por Pires *et al* [45].

Nessa abordagem, a principal diferença é que o SIFT Denso elimina a etapa de detecção de pontos e utiliza uma grade densa de pontos de interesse, por exemplo, todos os pixels da imagem ou uma grade que considera de 8 em 8 pixels. A Figura 3.8 ilustra um exemplo de grade densa em uma imagem onde cada círculo representa um ponto de interesse. Logo, cada ponto desta grade é descrito da mesma maneira que o SIFT por meio do gradiente da região vizinha.

A grande vantagem do SIFT Denso é que os pontos de interesse cobrem toda a imagem, pois as etapas de seleção e eliminação de pontos de interesse do SIFT não são executadas. Dessa forma, o número de pontos depende exclusivamente da grade utilizada na imagem. Por exemplo, se todos os pixels forem considerados como pontos de interesse, para uma imagem com $w \times h$ pixels são obtidos $M = w \times h$ pontos de interesse.

3.4 Histograma de Palavras Visuais (Bag of Visual Words - BoVW)

O histograma de palavras ou *Bag-of-Words (BoW)* é um algoritmo utilizado na área do reconhecimento de textos. Esta técnica extrai um histograma das ocorrências de palavras-chaves em um texto. Palavras-chaves são um conjunto de palavras comumente utilizadas para distinguir e discriminar o gênero do texto, tais como: científico, romântico, poético e etc.

Em 2004, Csurkan [15] propôs uma técnica baseada no BoW para realizar a categorização de imagens, mantendo a simplicidade e eficiência computacional do BoW. Esta técnica ficou

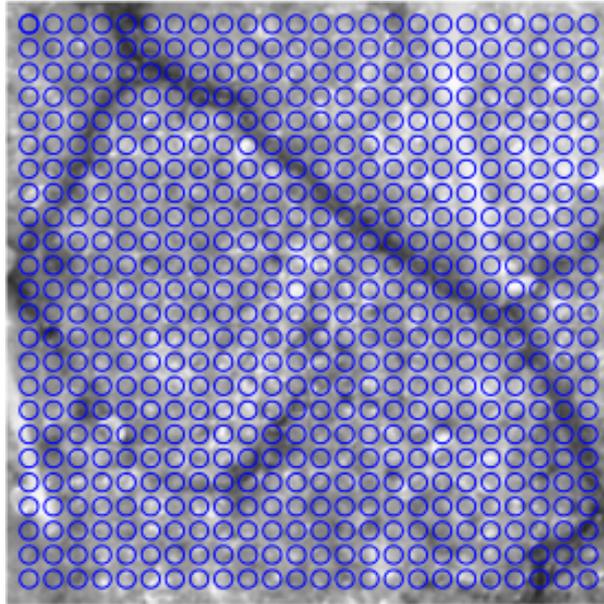


Figura 3.8: Exemplo de grade utilizada no SIFT Denso. Cada círculo representa um ponto de interesse na grade densa.

comumente conhecida como histograma de palavras visuais do inglês (*Bag-of-Visual-Words - BoVW*). Ela tem como base a obtenção de um conjunto de descritores, a criação de um vocabulário de palavras visuais utilizando técnicas de agrupamento e a construção de um histograma das ocorrências de palavras visuais na imagem.

A Figura 3.9 apresenta o fluxo de passos seguidos pelo BoVW. Em (a) é possível perceber o conjunto de treinamento do qual são calculados os descritores pelo SIFT ou pelo SIFT Denso, em (b) é realizada a construção do dicionário visual utilizando técnicas de agrupamento ou seleção aleatória de palavras. Depois, cada ponto de interesse é atribuído a palavra visual mais próxima em um passo conhecido como quantização. Por fim, temos um vetor de características, que representa o histograma das palavras visuais selecionadas em cada imagem. Nas seções seguintes são apresentadas cada uma das fases do BoVW em detalhes.

3.4.1 Detecção e Descrição de Pontos de Interesse

Em imagens, os pontos de interesse são pontos que se destacam e que podem ser utilizados em tarefas de reconhecimento. Estes pontos são regiões onde existe uma variação de intensidade dos *pixels* da vizinhança em diversas direções. A detecção e descrição desses pontos compreende a primeira e mais importante etapa para do BoVW, pois a representação final das imagens será definida de acordo com o conjunto de pontos de interesse detectados.

No mundo real, objetos apresentam-se em diferentes orientações, escalas e iluminações, o que dificulta a categorização das imagens e pode prejudicar os resultados. Para contornar esse problema são obtidos descritores que sejam invariantes a essas modificações, através de descritores locais como SIFT (*Scale Invariant Feature Transform*) [40] ou o SURF (*Speeded-Up Robust Feature*) [5]. Eles retornam um conjunto de descritores R , conforme Equação 3.12, que mapeia unicamente pontos de interesse dentro da imagem I e indicam a direção das variações que ocorrem em torno do ponto.

$$R_I = [r_1, r_2, \dots, r_M]^T \in \mathfrak{R}^{M \times G}, \quad (3.12)$$

onde r_i representa o descritor do ponto i , M representa a quantidade de pontos de interesse encontrados na imagem e G representa a dimensão dos descritores de cada ponto. A quanti-

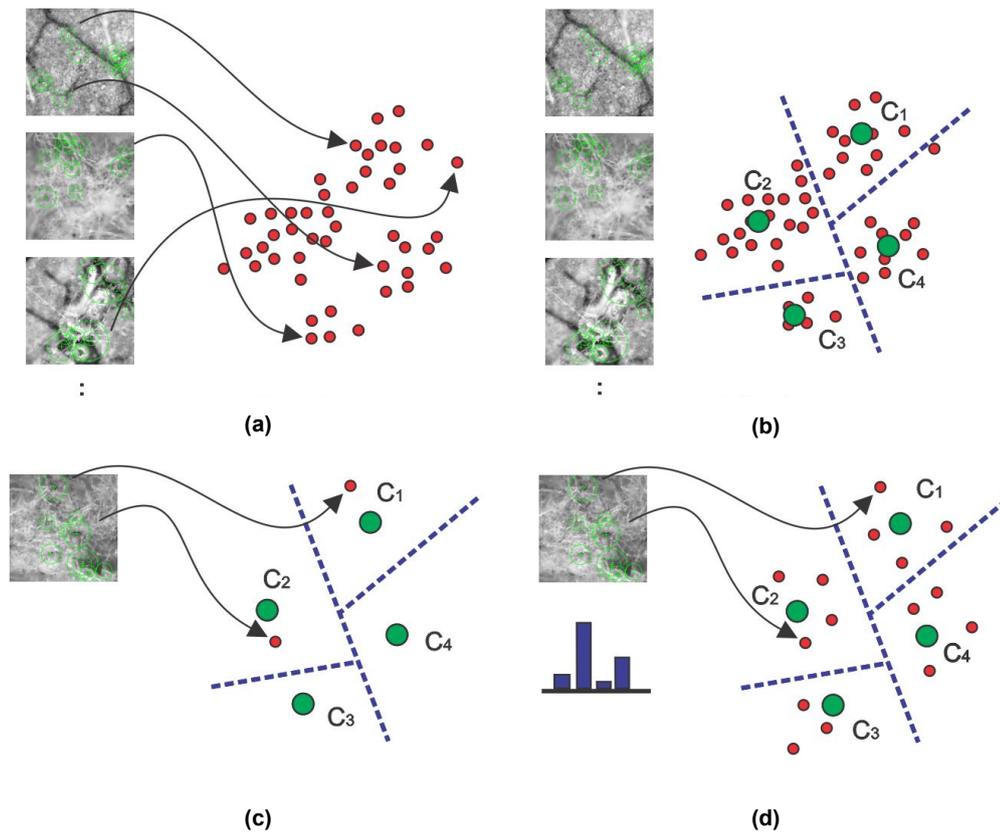


Figura 3.9: Passos do algoritmo BoVW (a) Corresponde a detecção e descrição de pontos de interesse. (b) Corresponde a criação do vocabulário visual. (c) Apresenta a contagem de palavras visuais. (d) Corresponde a criação do histograma. Imagem adaptada de [59].

dade de pontos de interesse M varia de imagem para imagem. Na Figura 3.10 são mostrados os descritores de pontos obtidos a partir do SIFT.

3.4.2 Criação do Vocabulário Visual

A construção de um vocabulário visual é realizada através da aplicação de um método de agrupamento para criar grupos com os descritores extraídos de um conjunto de imagens de treinamento. Para isso um descritor local é aplicado em cada imagem, conforme Equação 3.13:

$$D = [R_1, R_2, \dots, R_n], \quad (3.13)$$

onde n representa a quantidade de imagens de treinamento.

Segundo Fonseca [22], o algoritmo *K-means* tem se mostrado uma ótima alternativa para realizar o agrupamento de descritores e a localização de palavras visuais, por ser simples, intuitivo, e possuir uma complexidade linear. O *K-means* é aplicado sobre o conjunto de descritores D para obter um conjunto K de centroides, conforme Equação 3.14:

$$C = \text{K-means}(D), \quad (3.14)$$

O centroide de cada grupo corresponde a média dos valores de todos descritores pertencentes aquele grupo. No *K-means* a princípio os K centroides são definidos aleatoriamente, em seguida cada descritor do conjunto de entrada é associado ao centroide mais próximo e por

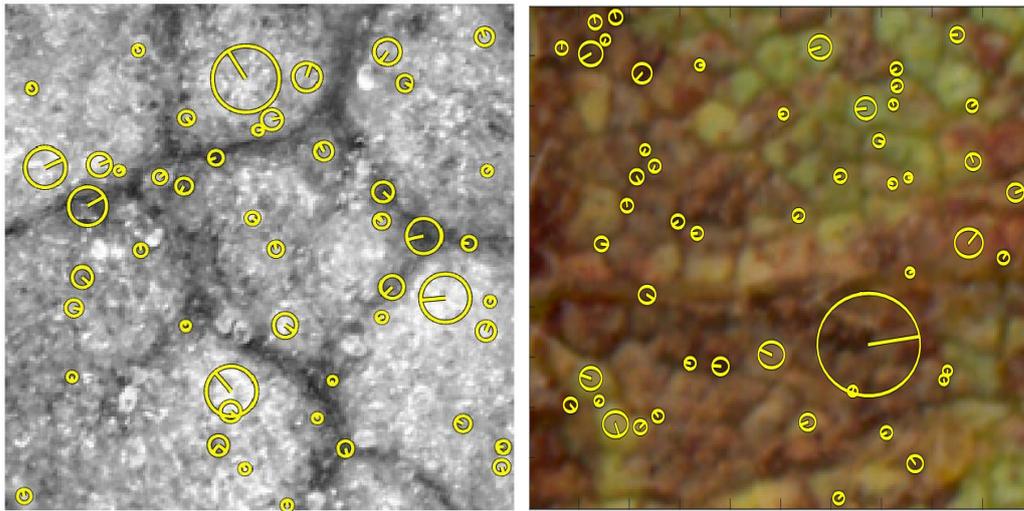


Figura 3.10: Descritores obtidos a partir do SIFT.

fim, cada centroide tem seu valor atualizado pela média de todos os descritores associados ao seu grupo. Esse processo é executado até que o centroide tenha seu valor estabilizado ou o limite de iterações seja atingido.

Cada centroide $c_i \in C$ corresponde a uma palavra visual e o conjunto das palavras visuais C formam o vocabulário de palavras visuais. Dessa forma, é possível quantificar a frequência de palavras visuais presentes em uma determinada imagem. Uma das desvantagens do K -means está justamente relacionada ao parâmetro K : um valor para K grande ou pequeno demais pode gerar um agrupamento de baixa qualidade. Além disso, um determinado ponto com valores altos demais pode causar uma grande alteração no centro de gravidade dos centroides, podendo ocasionar também um agrupamento de baixa qualidade.

A Figura 3.11, ilustra o resultado do agrupamento realizado utilizando o K -means sobre um dado conjunto de descritores gerando 5 centroides que compõe o vocabulário de palavras visuais.

3.4.3 Construção de Histograma

Cada imagem utilizada na etapa de formação do vocabulário, assim como aquelas que serão utilizadas durante o processo de classificação, são representadas na forma de um histograma. Dada uma imagem I e um vocabulário composto pelos centroides $C = c_1, c_2, \dots, c_k$, em que c_i denota uma palavra visual do vocabulário e K corresponde ao tamanho do vocabulário, a nova representação de I é dada pela frequência de cada palavra visual c_i em I .

Descritores são extraídos de I e, então cada ponto de interesse é rotulado pela palavra visual que apresentar a menor distância Euclidiana entre seu descritor local r_i e o centroide de cada palavra do vocabulário visual C . A palavra c_i mais próxima terá seu valor incrementado em sua respectiva posição no histograma.

A dimensão do histograma depende unicamente de K , pois este é a quantidade de palavras visuais do vocabulário.

Na Figura 3.12(a) é apresentado o processo de rotulação dos pontos de interesse pelo centroide mais próximo com base na distância Euclidiana. Na Figura 3.12(b) é realizada a quantização e geração do histograma de característica final da imagem.

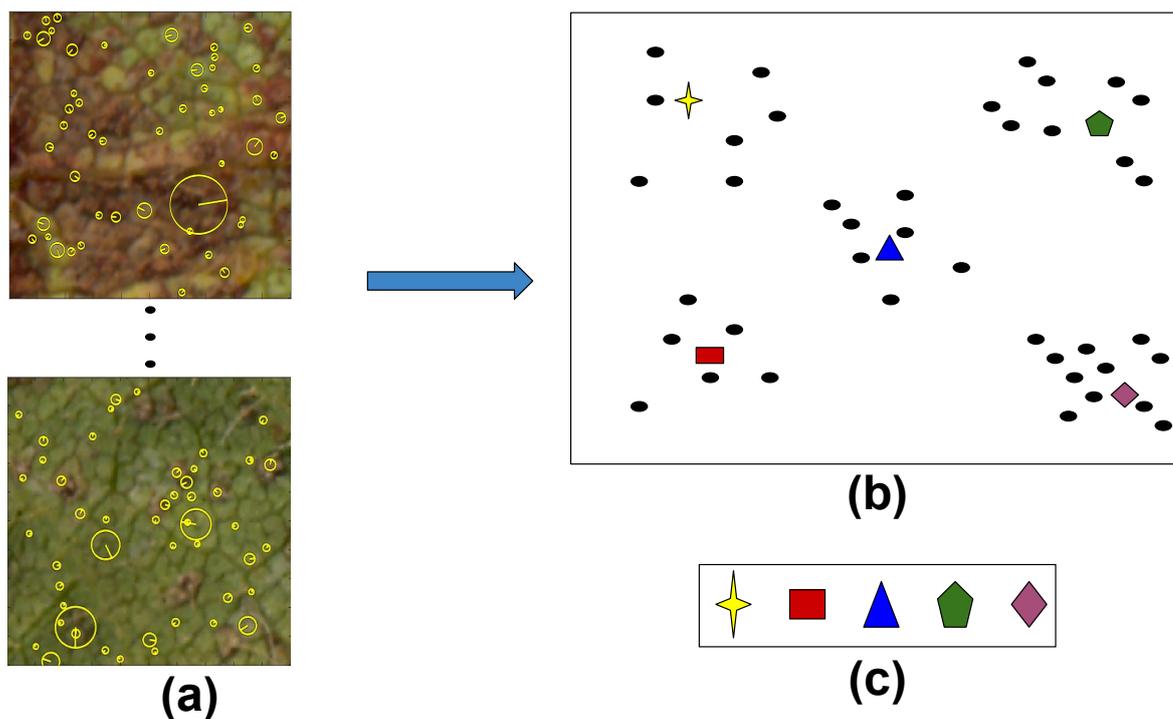


Figura 3.11: Construção do vocabulário visual a partir do *K-means*. (a) Conjunto de descritores extraídos, (b) Construção do vocabulário visual com o *K-means*, (c) Palavras visuais que compõem o vocabulário.

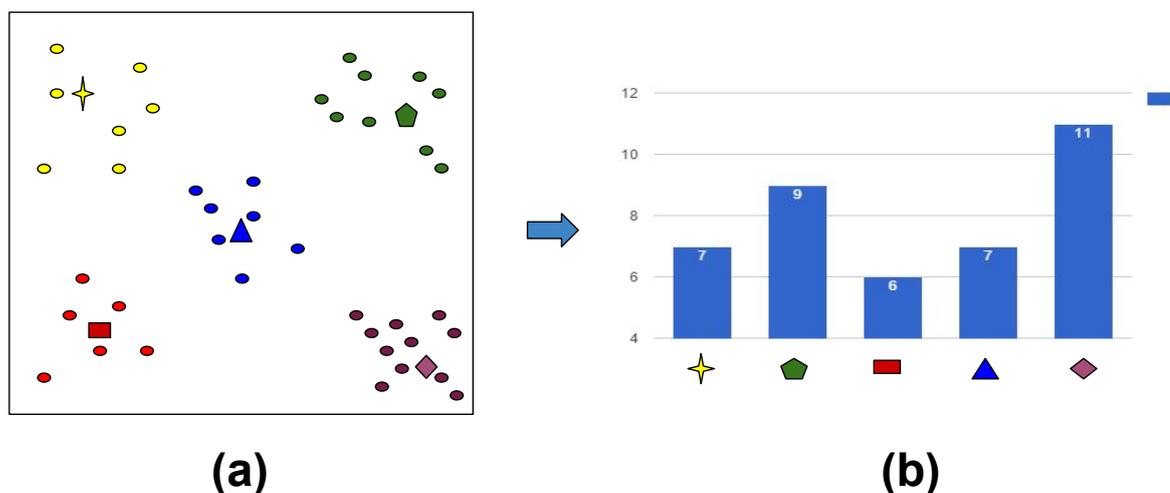


Figura 3.12: Construção do histograma da imagem, (a) Pontos de interesse rotulados pela palavra visual com menor distância Euclidiana. (b) Construção do histograma que representa o descritor final da imagem.

3.5 Vetores de Descritores Localmente Agregados (Vector of Locally Aggregated Descriptors - VLAD)

Com os avanços das pesquisas em classificação de imagens pelo método BoVW, foram identificados alguns problemas que podem gerar resultados inconclusivos ou mesmo incorretos na classificação. Dentre os problemas podemos citar a perda de informação local. O BoVW utiliza a chamada atribuição rígida [49], em que um descritor é associado a uma única

palavra visual, incrementado a contagem da mesma no histograma. Essa forma de rotulação pode gerar perda de informações locais no espaço de características pois não leva em consideração se a localização do descritor está em uma região de fronteira entre duas palavras visuais, se está muito distante de todas as palavras visuais (ruído), etc.

Jégou *et al* [32] propuseram o VLAD, que é uma representação vetorial de uma imagem que agrega descritores com base na sua localidade. O critério consiste na distância em que um descritor está da palavra visual mais próxima, com base no método conhecido como vizinho mais próximo ou (*Nearest Neighbors - NN*) [31]. Conseqüentemente é obtido um descritor de características de tamanho fixo, que pode então ser utilizado para a classificação da imagem.

Consideremos uma imagem I , em que foi obtido um conjunto R_I contendo M descritores. Com base no BoVW, inicialmente é criado um vocabulário C de palavras visuais com o método *K-means*, onde K corresponde a quantidade de palavras visuais no vocabulário. Cada um dos M descritores é então associado a uma palavra visual, utilizando *Nearest Neighbors - NN*, conforme Equação 3.15:

$$c_{r_i} = NN(r_i, C) \quad (3.15)$$

onde c_{r_i} indica a palavra visual ao qual o descritor r_i foi rotulado.

A Figura 3.13 apresenta um exemplo de rotulação através da Equação 3.15. Esta figura toma como base cinco palavras visuais existentes na Figura 3.12(a) representadas por c_1 , c_2 , c_3 , c_4 e c_5 , e para efeito de simplificação na visualização foram considerados cinco descritores representados por r_1 , r_2 , r_3 , r_4 e r_5 . Dentre esses descritores, um é atribuído a palavra c_1 , três são atribuídos a c_2 , e um para c_5 .

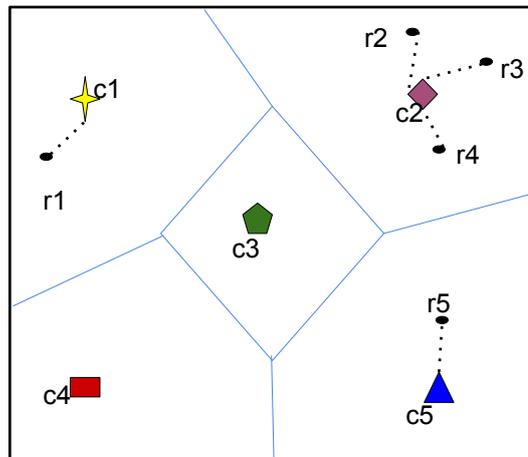


Figura 3.13: Rotulação dos descritores pela palavra visual mais próxima.

Assim que um descritor r_i é associado a palavra visual c_{r_i} mais próxima no vocabulário, é calculada a diferença entre os valores de ambos, conforme Equação 3.16:

$$v_i = r_i - c_{r_i} \quad (3.16)$$

tal que v_i corresponde ao vetor diferença do descritor r_i e sua palavra visual c_{r_i} .

Essa diferença é calculada para cada um dos M descritores pertencentes a imagem I associados a alguma palavra visual pertencente ao vocabulário C . Na Figura 3.14, é apresentado o vetor de diferença entre cada um dos M descritores e palavra visual a qual estão atribuídos. Nesta figura, para cada uma das 5 palavras visuais, são apresentadas as diferenças para os descritores que foram atribuídos a elas. A diferença é representada pelas setas em azul.

Podemos perceber que quanto mais próximo o descritor esta da palavra visual, menor será o vetor de diferença. Portanto esta informação local é mais rica quando comparada apenas com a atribuição rígida.

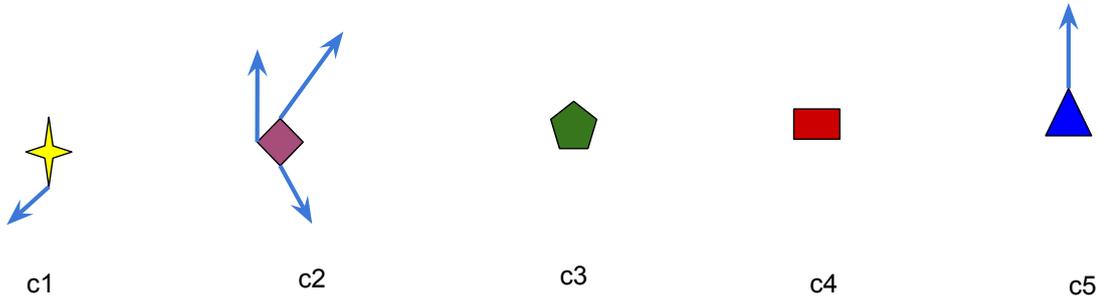


Figura 3.14: Cálculo da diferença entre $r_i - c_{r_i}$ para cada descritor rotulado por uma palavra visual.

Para sumarizar, em cada palavra visual os vetores v_i são acumulados conforme Equação 3.17:

$$S_{c_i} = \sum_{i=1}^M v_i, \text{ Se } c_{r_i} == c_i \quad (3.17)$$

tal que S_{c_i} representa a soma dos vetores de diferença atribuídos a palavra visual c_i . Na Figura 3.15 as palavras visuais são representadas por c1 a c5 e o somatório dos vetores diferença resulta em um vetor representado pelas setas em vermelho.

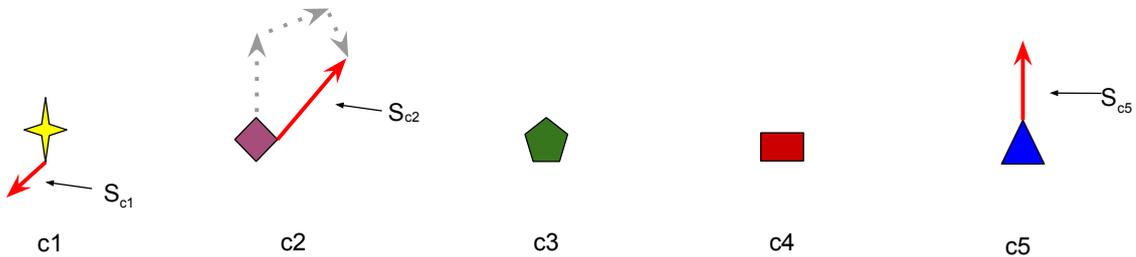


Figura 3.15: Somatório das diferenças de cada descritor r_i ao centroide c_{r_i} correspondente, para a construção do vetor v final.

Assim, o descritor de características final da imagem I é composto pela concatenação dos S_{c_i} de cada palavra visual:

$$\Phi_I = [S_{c_1}, S_{c_2}, \dots, S_{c_k}] \quad (3.18)$$

o vetor Φ_I possui tamanho fixo, contendo $K \times G$ elementos, sendo K o número de palavras visuais presentes no vocabulário e G o tamanho do descritor, no caso do SIFT temos $G = 128$.

Resultados experimentais demonstraram que excelentes resultados podem ser obtidos mesmo com um número, relativamente, pequeno de palavras visuais. Segundo Jégou *et al* [32], o valor ideal do parâmetro K para a tamanho dos vocabulários no VLAD estaria entre 16 e 256.

3.5.1 Normalizações

O VLAD foi inicialmente proposto com o objetivo de permitir a construção de vetores de características mais discriminativos com informações locais dos descritores. No entanto, mesmo essa forma de representação está sujeita a erros de acordo com as características da imagem, pois no processo de construção dos vetores podem existir valores artificialmente grandes, como por exemplo: estruturas repetidas no fundo de uma imagem [3]. Para tratar esse problema foram propostas formas de normalização que tem como objetivo excluir valores muito altos ou baixos que poderiam afetar negativamente o desempenho da classificação.

Uma das primeiras propostas de normalização é a chamada *component-wise normalization*, onde cada vetor S_{c_i} é normalizado pela sua norma, conforme Equação 3.19:

$$S_{c_i} = \frac{S_{c_i}}{\|S_{c_i}\|_2}, \forall c_i \in C \quad (3.19)$$

onde:

$$\|S_{c_i}\|_2 = \sqrt{\sum_{j=1}^G S_{c_i}(j)^2} \quad (3.20)$$

Uma segunda forma de normalização proposta no VLAD é com base na raiz quadrada ou *square-rooting*. Nesta forma de normalização cada elemento do vetor S_{c_i} é normalizado pela sua raiz quadrada, conforme Equação 3.21:

$$S_{c_i}(j) = \text{sign}(S_{c_i}(j)) \times \sqrt{|S_{c_i}(j)|}, j = 1, 2, \dots, G \quad (3.21)$$

onde $\text{sign}(Z)$ representa o sinal (positivo, negativo) de Z .

Por fim, foi proposta uma terceira forma de normalização, a chamada *componet-wise mass normalization*. Nessa abordagem cada vetor S_{c_i} é dividido pela soma do total de descritores atribuídos a ele, conforme Equação 3.22:

$$S_{c_i} = \frac{S_{c_i}}{m_{c_i}} \quad (3.22)$$

onde:

$$m_{c_i} = \sum_{i=1}^M \begin{cases} 1, & \text{se } c_{r_i} = c_i \\ 0, & \text{caso contrário} \end{cases} \quad (3.23)$$

As normalizações são implementadas como uma função opcional no VLAD podendo melhorar os resultados da classificação através de um vetor mais compacto. Outra vantagem das normalizações é que podem ser utilizadas em conjunto, podendo simplificar ainda mais o vetor VLAD.

Abordagens Propostas

Com os avanços das pesquisas em classificação de imagens pelo método BoVW, foram propostas variações visando a construção de descritores mais robustos e que fornecessem características mais relevantes. Uma dessas propostas é o VLAD [32], que utiliza vetores para agregar os descritores de uma imagem como base em critérios de localidade. Porém essa solução, tal qual o BoVW, utiliza uma atribuição rígida em que um descritor é atribuído a um único centroide, fornecendo descritores com perda de informação local. Por exemplo, considere duas situações: i) um descritor está a uma distância semelhante de dois centroides; ii) um descritor está a uma distância relativamente grande de todos os centroides. Na primeira situação, tendo como base a atribuição rígida, o descritor seria atribuído exclusivamente a um centroide mesmo estando a uma distância semelhante de ambos. Já na segunda situação, o descritor seria associado ao centroide com a menor distância, mesmo que essa distância caracterize um ruído.

Uma solução para o problema da atribuição rígida foi proposta por Gemert *et al.* [63] que utilizaram a atribuição leve ao vocabulário visual através da ambiguidade, permitindo que um descritor seja atribuído a mais de um centroide. Neste trabalho foram realizadas comparações entre quatro formas de atribuição: rígida, núcleo Gaussiano, ambiguidade pela incerteza da palavra visual e ambiguidade pela plausibilidade da palavra visual. A ambiguidade pela incerteza da palavra visual indica que um descritor pode ser atribuído para mais de uma palavra visual. Já a ambiguidade pela plausibilidade da palavra visual assume que um determinado descritor pode não estar perto suficiente para justificar a representação por qualquer palavra visual.

Com base no VLAD e nos estudos realizados por Gemert *et al.* [63], propomos um novo descritor de imagens que permite realizar a atribuição de um descritor aos centroides de forma proporcional a sua distância. O peso da atribuição é calculado com base na distância entre o descritor e os centroides, de tal forma que, quanto mais distante um descritor estiver de um centroide, menor é o peso do descritor para este centroide. Seguindo a ideia apresentada acima, foram propostas três abordagens. A primeira abordagem atribui um descritor a todos os centroides ponderado pelos pesos. A segunda abordagem atribui um descritor aos P centroides mais próximos. Por fim, a terceira abordagem atribui um descritor a um número variável de centroides com base na distribuição de distâncias.

As abordagens propostas resumem-se nas seguinte etapas:

1. Detecção e descrição dos pontos de interesse por um descritor local (por exemplo, o SIFT).
2. Criação do vocabulário de palavras visuais. Nessa etapa foi utilizado o *K-means* e um conjunto de imagens de treinamento.
3. Rotulação dos descritores para as palavras visuais. Nessa etapa são calculados os pesos e a atribuição leve, permitindo que um mesmo descritor seja rotulado para mais de uma palavra visual.
4. Construção do VLAD utilizando as rotulações e pesos calculados na etapa anterior.

A Figura 4.1 ilustra o processo de cálculo do peso de um descritor r_1 para um conjunto de 5 centroides (c_1, c_2, c_3, c_4 e c_5). Em (a) são ilustrados os descritores e os centroides sobre uma imagem de folha de soja doente; em (b) é apresentado o calculo da distância do descritor r_1 para todos os centroides e em (c) é temos o calculo do peso do descritor r_1 para cada um dos centroides e em (d) é apresentada a construção dos vetores VLAD de forma que quanto mais próximo o descritor r_1 estiver do centroide, maior é o peso que ele exerce e portanto mais espessa é a seta indicando o vetor.

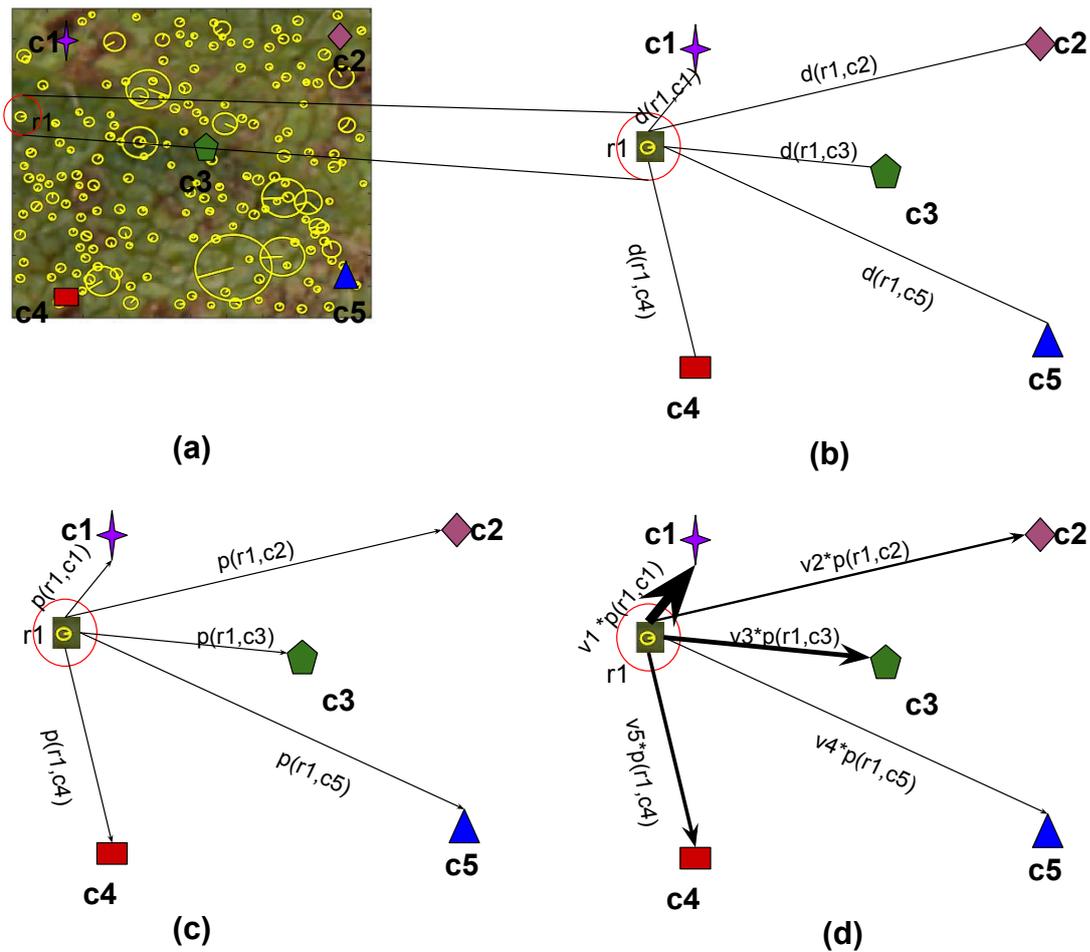


Figura 4.1: Etapas seguidas nas abordagens propostas para o calculo do peso de um descritor para os centroides. (a) Obtenção dos descritores e centroides; (b) Calculo da distância do descritor r_1 para os centroides, (c) O calculo do peso do descritor r_1 (d) Aplicação dos pesos aos vetores VLAD.

4.1 Atribuição Leve e Pesos

O cálculo de peso entre um descritor e um centroide leva em consideração a distância envolvida entre eles, tal que, quanto maior a distância entre um descritor r_i e um centroide c_j , menor o peso que o descritor r_i terá para o centroide c_j . Desta forma, podemos considerar pesos parciais para descritores em regiões de fronteira e desconsiderar descritores que estão muito distantes do centroide, pois o peso seria muito baixo. O cálculo do peso dos descritores é realizado da seguinte forma:

1. Primeiro passo é calcular a distância entre os descritores de uma imagem e os centroides, assim como no BoVW. A distância pode ser calculada através de diversas métricas, entre as mais comuns está a distância Euclidiana, conforme Equação 4.1:

$$dist(r_i, c_j) = \|r_i - c_j\| \quad (4.1)$$

onde r_i e c_j são, respectivamente, um descritor e um centroide.

Visando tornar as distâncias entre os descritores e os centroides inversamente proporcionais, a distância é normalizada através da Equação 4.2:

$$dist(r_i, c_j) = \frac{1}{dist(r_i, c_j)} \quad (4.2)$$

2. Em seguida, para cada descritor r_i é calculada a soma das distâncias para todos os centroides c_j , dada pela Equação 4.3:

$$s(r_i) = \sum_{j=1}^K dist(r_i, c_j) \quad (4.3)$$

3. Por fim, para cada distância $dist(r_i, c_j)$ é calculada a probabilidade do descritor r_i ser atribuído ao centroide c_j , conforme a Equação 4.4:

$$p(r_i, c_j) = \frac{dist(r_i, c_j)}{s(r_i)} \quad (4.4)$$

Após a realização dos passos acima, cada descritor possui um peso para cada centroide levando em consideração a distância entre eles. A próxima etapa, seguindo o VLAD, consiste em aplicar os pesos nas diferenças entre os descritores e os centroides, conforme Equação 4.5:

$$\sigma_{r_i, c_j} = p(r_i, c_j) * (r_i - c_j), \quad (4.5)$$

onde σ_{r_i, c_j} representa o vetor diferença entre um descritor e o seu centroide normalizado pelo peso. Em seguida é necessário somar os vetores para obter a contribuição da palavra visual conforme a Equação 4.6:

$$S_{c_j} = \sum_{i=1}^M \sigma_{r_i, c_j} \quad (4.6)$$

tal que S_{c_i} contém a soma dos vetores de diferença atribuídos a palavra visual normalizado pelo peso que os descritores exercem.

Nas seções abaixo, as três abordagens propostas nesse trabalho são descritas em detalhes. Em cada uma das abordagens, os pesos descritos acima são utilizados para ponderar a

atribuição dos descritores aos centroides e para definir o número de centroides em que cada descritor é atribuído.

4.2 VLAD-F: Vector of Locally Aggregated Descriptors - Full

A primeira abordagem proposta atribui cada descritor a todos os centroides na construção dos vetores com o VLAD. A principal vantagem desta abordagem é fornecer uma grande quantidade de informações para o cálculo dos vetores pelo VLAD pois faz uso de todos os descritores e centroides.

Similarmente ao VLAD tradicional, em sua primeira etapa, é aplicado um descritor de pontos locais (por exemplo, o SIFT) para obter os descritores; a segunda etapa é a criação do vocabulário de palavras visuais pelo *K-means*. A terceira etapa é responsável pela rotulação dos descritores para os centroides, nessa etapa é calculada as distâncias e os pesos $p(r_i, c_j)$ entre os descritores e centroides. Em seguida todos os descritores são rotulados para todos os centroides. Por fim, na quarta etapa são construídos os vetores VLAD utilizando os pesos dos descritores para normalizar os valores dos vetores.

A Figura 4.2 ilustra o processo de construção dos vetores VLAD utilizando a primeira abordagem que rotula um descritor para todos os centroides. Figura 4.2(a) representa aplicação do descritor local SIFT, enquanto que a Figura 4.2(b) ilustra o vocabulário de palavra visuais e apresenta a distância de um descritor r_1 para todos os centroides. A Figura 4.2(c) apresenta os pesos deste descritor para todos os centroides de tal forma que quanto mais perto do centroide, maior o peso (e.g., $p(r_1, c_1) > p(r_1, c_4)$). Por fim a Figura 4.2(d) apresenta os vetores normalizados pelos pesos associados, sendo possível notar que foi traçado um vetor direcional para cada centroide.

4.3 VLAD-PNC: Vector of Locally Aggregated Descriptors P-Nearest Centroids

A atribuição de todos os descritores para todos os centroides, como na primeira abordagem, tem como principal vantagem fornecer um grande conjunto de informações para o cálculo dos vetores VLAD. Todavia, essa alternativa pode recair nos mesmos problemas encontrados em uma atribuição rígida, pois descritores que estivessem muito distantes também são rotulados aos centroides e as informações adicionadas afetariam negativamente no desempenho. Uma alternativa é limitar o número de centroides em que um descritor pode ser rotulado.

O VLAD com P centroides mais próximos (VLAD-PNC) consiste em atribuir um determinado descritor a uma quantidade P de centroides mais próximos, e dessa forma, impedir que a rotulação atribua o descritor a centroides muito distantes. A Figura 4.3 ilustra o processo de construção dos vetores VLAD utilizando a segunda abordagem que rotula um descritor para um conjunto $P = 2$ de centroides. As Figuras 4.3(a), (b) e (c) apresentam os descritores, o vocabulário de palavra visuais e o peso do descritor r_1 para todos os centroides, respectivamente. Na segunda abordagem, a Figura 4.3(d) apresenta os vetores normalizados pelos pesos em que nota-se a rotulação do descritor r_1 aos dois centroides com maior peso (c_1 e c_3).

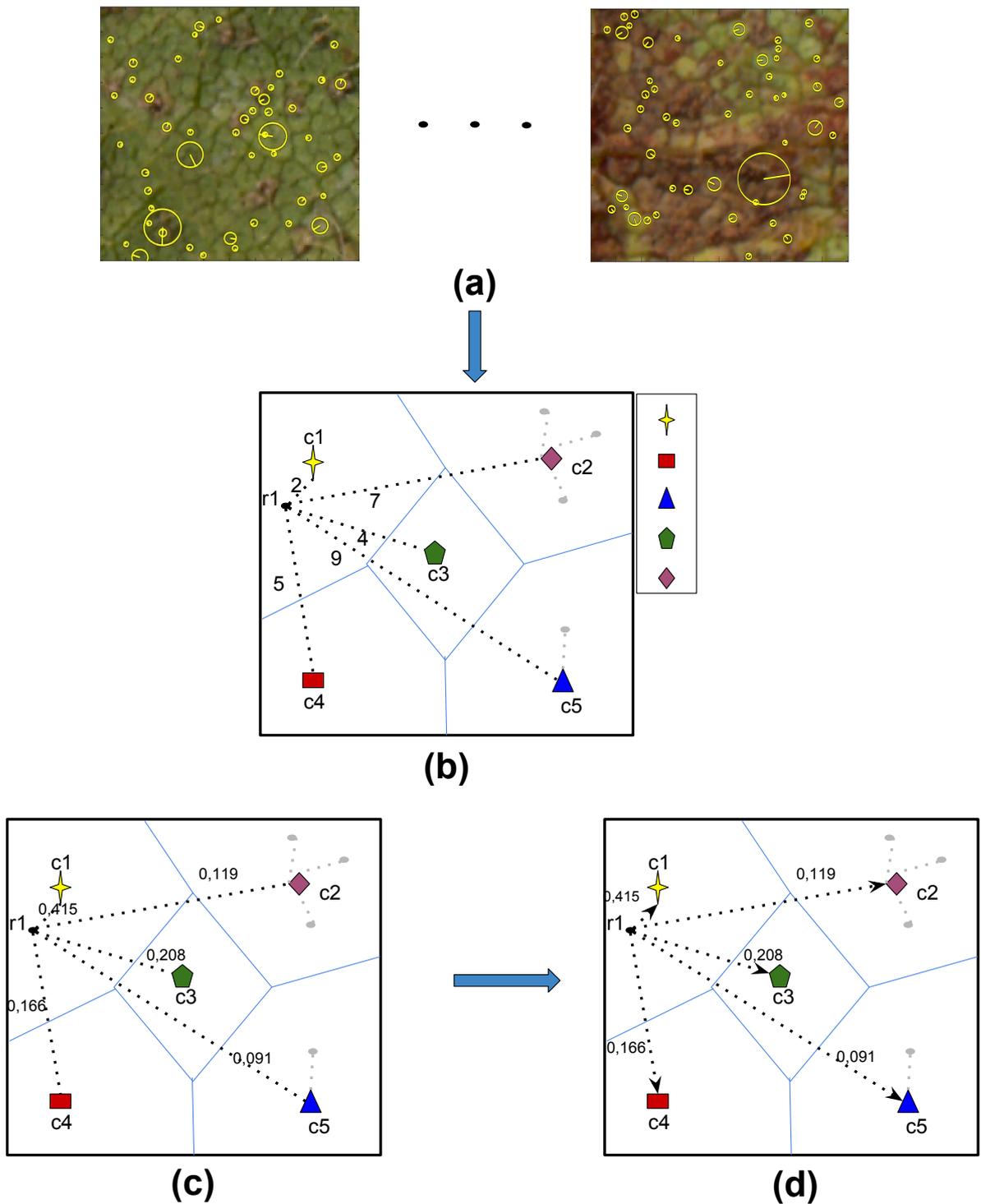


Figura 4.2: Primeira abordagem: Todos os descritores são atribuídos a todos os centroides.

4.4 VLAD-VNNC: Vector of Locally Aggregated Descriptors - Variable Number of Nearest Centroids

A utilização de um número P de centroides mais próximos traz como principal vantagem impedir a rotulação de descritores a centroides muito distantes. Porém, fixar um número de centroides mais próximos pode não ser adequado para todas as situações, por exemplo, quando um descritor está bem próximo de apenas um centroide e distante dos demais.

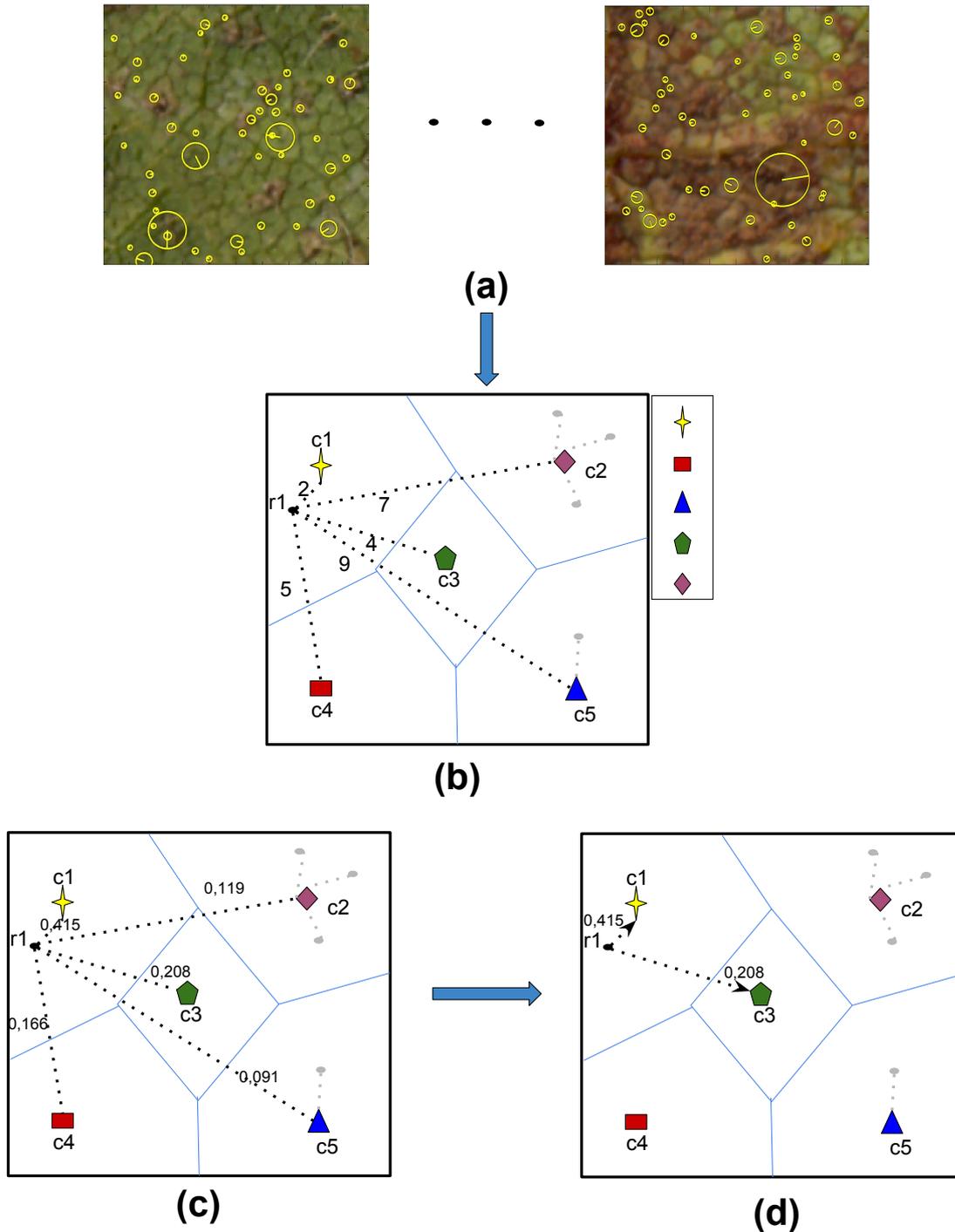


Figura 4.3: Segunda abordagem: atribui um descritor a um conjunto P de centroides com maior peso. Nessa imagem temos $P = 2$, assim cada descritor é rotulado a dois centroides.

Para a terceira abordagem, considere as distâncias $dist(r_i, c_j)$ de um descritor r_i para todos os centroides. O descritor r_i é atribuído a um centroide c_j se a sua distância for pelo menos 80% da menor distância:

$$p(r_i, c_j) = \begin{cases} p(r_i, c_j), & \text{se } \minDist(r_i)/dist(r_i, c_j) \geq 0.8 \\ 0, & \text{caso contrário} \end{cases} \quad (4.7)$$

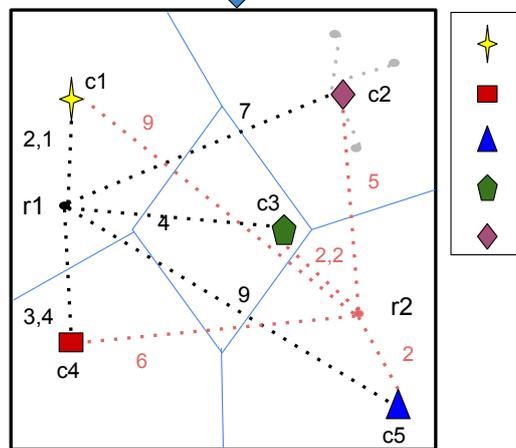
onde $\minDist(r_i) = \min_{c_j} dist(r_i, c_j)$ é a distância do centroide mais próximo. A principal vantagem dessa abordagem é atribuir um descritor a mais de um centroide desconsiderando centroi-

des muito distantes e permitir uma variação quanto a quantidade de centroides ao qual um descritor pode ser atribuído.

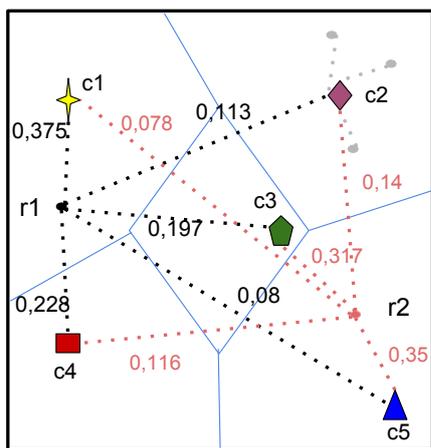
A Figura 4.4 ilustra o processo de construção dos vetores VLAD utilizando a terceira abordagem que rotula um descritor para um conjunto variável de centroides. A Figura 4.4(a) apresenta a aplicação do descritor local SIFT. Na Figura 4.4(b), é apresentado o vocabulário de palavra visuais e as distâncias de dois descritores, (r_1, r_2) , para todos os centroides obtidos. O peso dos dois descritores para os centroides é apresentado na Figura 4.4(c). Por fim, a Figura 4.4(d) apresenta os vetores construídos e sendo normalizados pelos pesos, onde nota-se que o descritor r_1 foi atribuído a apenas um centroides (c_1), enquanto que o descritor r_2 foi atribuído a dois (c_3 e c_5). Dessa forma, cada descritor é atribuído a um número variável de centroides que depende da distribuição das distâncias.



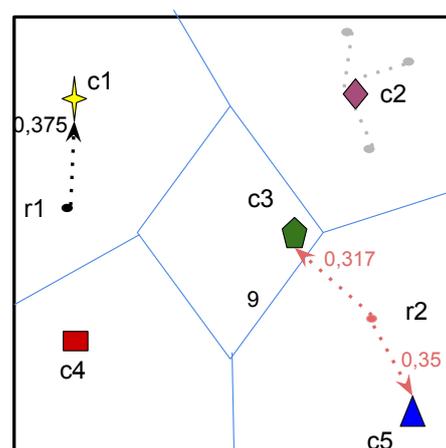
(a)



(b)



(c)



(d)

Figura 4.4: Terceira abordagem: os descritores são atribuído a um número variável de centros de acordo com o peso que eles tem para o centroide.

Experimentos e Resultados

Este capítulo descreve os experimentos e respectivos resultados obtidos pelas técnicas propostas utilizando o VLAD para a tarefa de reconhecimento de doenças foliares da cultura de soja. Para isso foram utilizadas duas bases de imagens descritas na Seção 5.1, contendo imagens de folhas de soja correspondentes as classes do problema. A Seção 5.2 descreve os experimentos realizados. Neles foram utilizados dois descritores locais: SIFT [40] e o SIFT Denso [38], juntamente com as abordagens propostas no Capítulo 4. Os resultados obtidos pelas técnicas foram comparados aos do BoVW e o VLAD, com os descritores SIFT e SIFT Denso.

A Figura 5.1 ilustra cada uma das etapas que foram seguidas para a realização dos experimentos:

5.1 Base de Imagens de Soja

5.1.1 Soja Scanner

A primeira base de imagens foi construída como parte de um experimento junto ao departamento de Fitopatologia da Universidade Federal de Grande Dourados (UFGD), Brasil [53]. A colheita avaliada foi de soja (*Glycine max (L.) Merr*), em que a densidade dos campos de soja foi aproximadamente 300.000 plantas por hectare.

As folhas foram coletadas aleatoriamente em três estágios diferentes: V4 - (quarto trifólio), V5 - (quinto trifólio) e R1 - (estágio de florescência). Nas fases V4 e V5, nove plantas foram coletadas aleatoriamente por parcela para avaliação de doenças foliares, especialmente aquelas relacionadas aos fungos. Na fase R1, outras nove plantas foram coletadas para avaliação. O material coletado dividiu-se nos trifólios da fase de crescimento. Para a região Centro-Oeste, duas classes de doenças são mais comumente encontradas: a ferrugem e o míldio.

Com o apoio de um especialista, durante a coleta da doença ferrugem foram criados grupos classificados de acordo com as características de cor apresentadas nas lesões das folhas: RB e TAN. As lesões RB são caracterizadas por uma coloração marrom avermelhada e são consideradas um tipo de lesão mais resistente quando comparadas com as lesões TAN. Por outro lado as lesões de míldio são caracterizadas por manchas de coloração verde-clara até

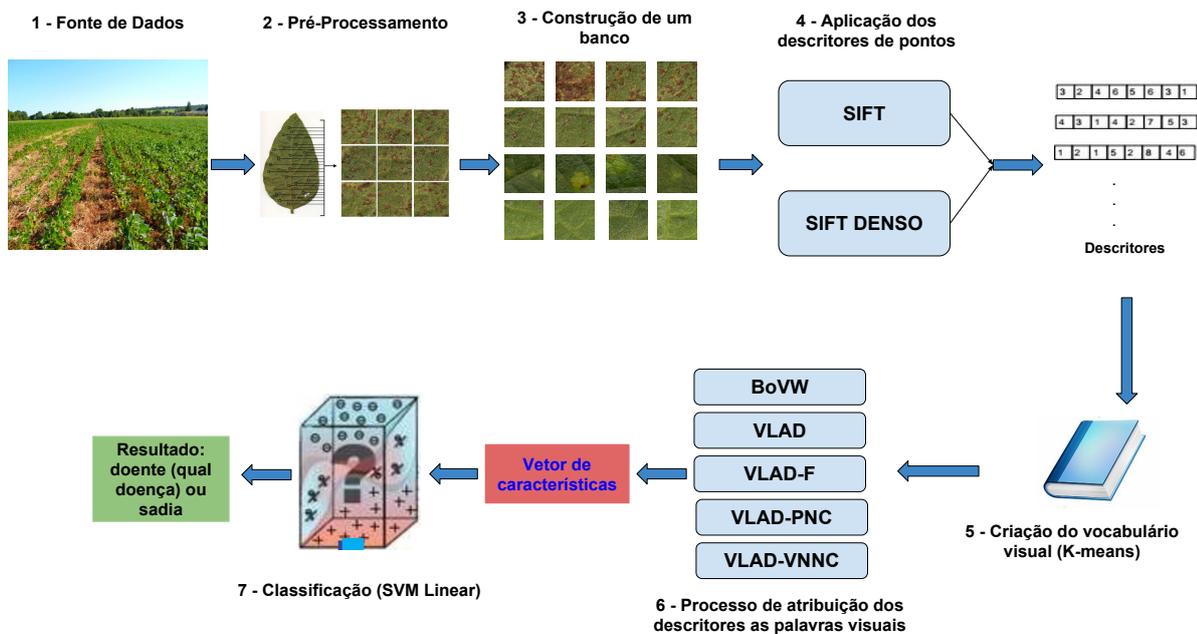
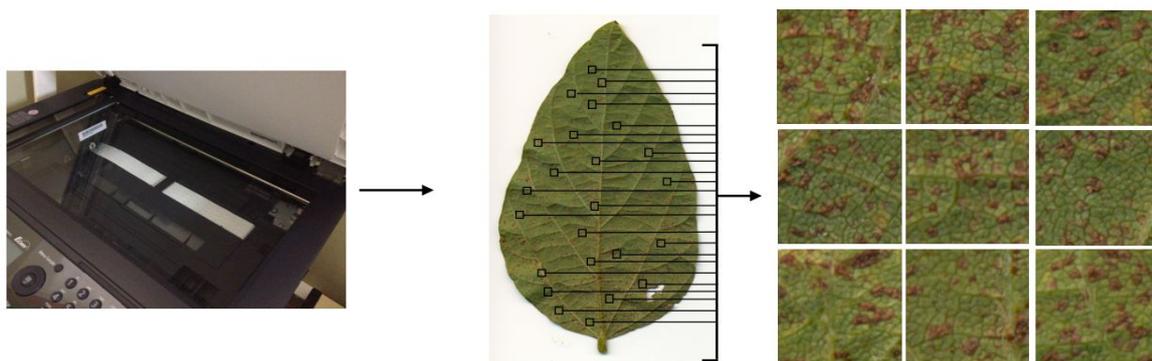


Figura 5.1: Processo de realização dos experimentos. 1) Fonte de dados, 2) Pré-Processamento feito com o auxílio de um especialista da área, 3) Construção de um banco de imagens com amostras das folhas doentes e saudáveis, 4) Aplicação de um dos descritores de pontos locais estudados (SIFT, SIFT Denso), 5) Construção do vocabulário de palavras visuais a partir dos descritores da imagens, 6) Processo de atribuição dos descritores as palavras visuais de acordo com a técnica selecionada e 7) Classificação por meio do SVM Linear.

amareladas, localizadas na face da folha.

A Figura 5.2 apresenta o processo de construção do banco a partir das folhas coletadas. A abordagem é composta por 3 estágios principais, depois da coleta das folhas com o auxílio de especialista: Figura 5.2(a) mostra o processo de digitalização das folhas, na Figura 5.2(b) temos o processo de seleção das amostras com base nas texturas presentes nas folhas, e na Figura 5.2(c) vemos o conjunto de amostras de textura.



(a) Digitalização das folhas. (b) Seleção das Amostras de textura. (c) Conjunto de amostras de textura.

Figura 5.2: Procedimento de obtenção das imagens utilizado nesse estudo. O banco é composto por 4 classes de imagens que representam o problema de classificação de doenças foliares da cultura de soja. Adaptada de [53].

No estudo foi utilizado um scanner de mesa HP Scanjet 1200, com resolução de 1200 dpi. Foram geradas imagens em formato *tiff* que posteriormente foram convertidas para o formato *png* para o experimento, que permite a representação das características das folhas

em uma paleta de 256 cores. Para cada folha, foram coletadas nove amostras, digitalizadas em janelas de tamanho 200×200 pixels de manchas específicas, identificadas por especialistas como doenças. Para a composição do banco de imagens foram selecionadas 300 janelas de cada doença (ferrugem RB, ferrugem TAN e míldio), totalizando 900 imagens. Além disso, foram coletadas outras 300 janelas de folhas sem nenhuma lesão ou ferimento, para compor o conjunto de plantas saudáveis. Portanto, o conjunto total de imagens é composto por 1200 amostras de imagens, divididas em 4 classes, conforme Figura 5.3.

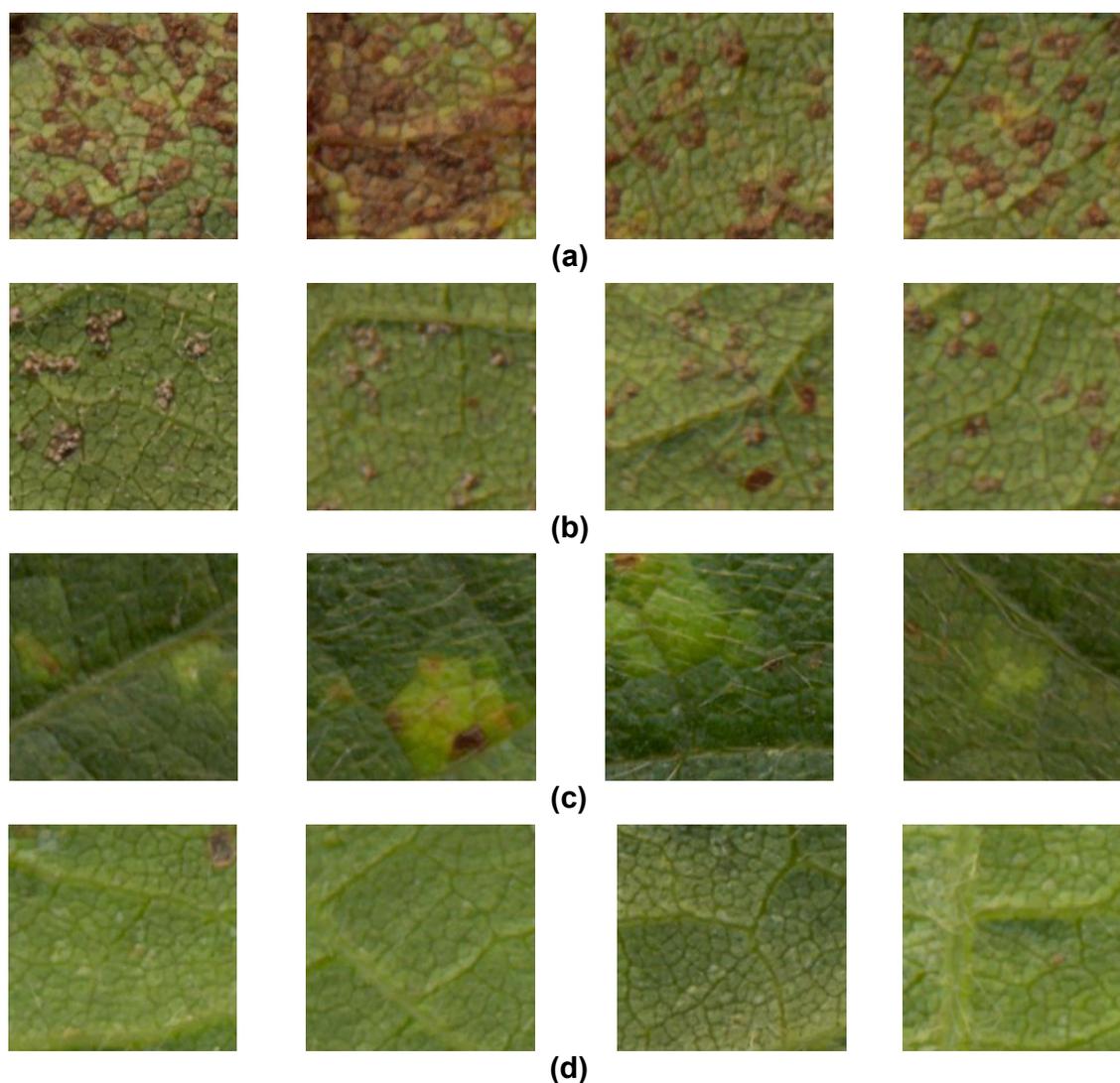


Figura 5.3: Conjunto de imagens das classes do problema de identificação de doenças foliares da cultura de soja. (a) Classe de folhas com a doença ferrugem asiática RB, (b) Classe de folhas com a doença da ferrugem asiática TAN, (c) Classe de folhas com a doença míldio e (d) Classe de folhas saudáveis.

5.1.2 Plant Village

A PlantVillage [30] é uma plataforma on-line dedicada ao estudo e pesquisa de doenças de plantas, disponível em www.plantvillage.org. Ela foi construída por David P. Hughes, *University Park PA, USA* e Marcel Salathé, *School of Computer and Communication Sciences, EPFL, Switzerland*. Atualmente, é mantida por diversos pesquisadores e profissionais das áreas de agricultura e tecnologia. O funcionamento é semelhante ao de um fórum, em que os usuários podem interagir entre si através de perguntas e respostas.

A PlantVillage fornece informações sobre mais de 150 culturas, e mais de 1800 doenças totalizando 87280 imagens, sendo o conteúdo validado por especialistas em patologias de plantas e por toda a comunidade presente. Todas imagens presentes no banco de dados do PlantVillage foram obtidas em estações de pesquisa associadas ao *Land Grant Universities*, USA, que reúne universidades como: *Penn State*, *Florida State*, *Cornell*, entre outras.

A maioria das imagens foram coletadas por dois técnicos em experimentos em campo de culturas infectadas com as doenças. Todas as imagens foram tiradas sob a luz do sol, e intencionalmente se procurou um intervalo de condições tais como o usuário final (produtor com um *smartphone*). Para cada folha, foram coletadas entre 4 e 7 imagens utilizando uma câmera digital padrão (Sony DSC - Rx100 / 13 20,2 megapixels) com disparo em modo automático.

Para a cultura de soja o banco conta hoje com 12674 imagens, sendo 6235 de folhas saudáveis, 851 de folhas doentes com Míldio (*Peronospora manshurica*), 2023 de folhas doentes com Mancha Olho de Rã (*Frogeye leaf spot*), 3565 de folhas doentes com Mancha Parda (*Septoria Leaf Blight*).

A Figura 5.4 ilustra algumas imagens presentes na base Plant Village, sem nenhum tipo de pré-processamento. A Figura 5.4(a) corresponde a uma folha doente de míldio com tamanho de 800×658 pixels, em (b) está representada uma amostra com a doença mancha olho de rã de tamanho 800×533 pixels, (c) tem se uma folha com a doença da mancha parda de dimensões 533×800 pixels, e (d) apresenta uma imagem de folha saudável de tamanho 800×533 pixels.

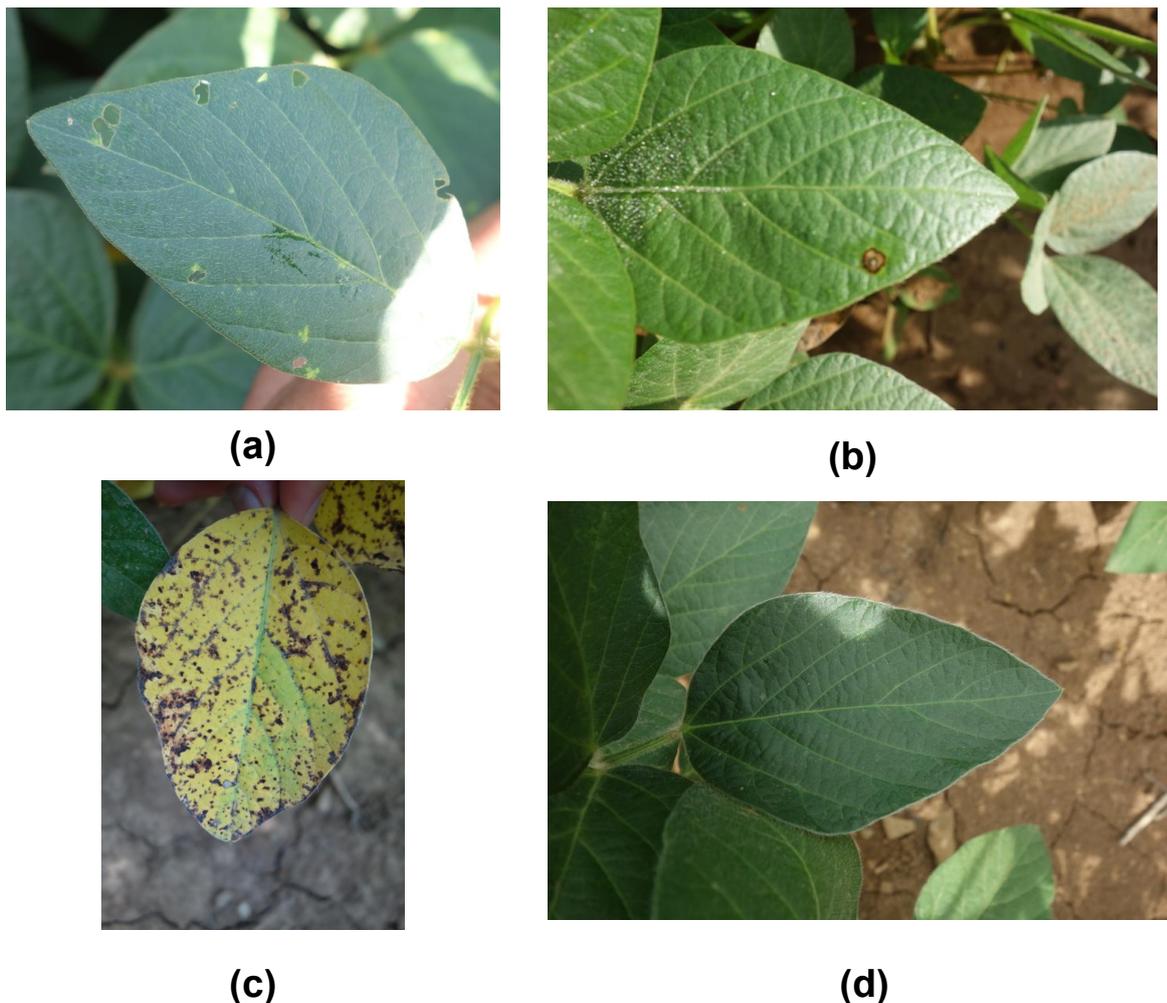


Figura 5.4: Conjunto de imagens da base Plant Village sem nenhum pré-processamento, disponível em www.plantvillage.org. Em (a) folha com a doença míldio, (b) folha com a doença mancha olho de rã, (c) folha com a doença da mancha parda, e (d) folha saudável.

As imagens presentes no banco foram submetidas a um pré-processamento com o apoio de especialistas da área, em que foram excluídas imagens pouco representativas e foram selecionados pontos das folhas correspondentes a características das doenças para a construção de um conjunto de amostras.

A partir desse pré-processamento foi obtido um conjunto de 7584 imagens de tamanho padrão 199×201 , sendo: 851 imagens de Míldio, 2023 imagens de Mancha Olho de Rã, 3565 imagens de Mancha Parda e 1145 imagens de folhas saudáveis. A Figura 5.5 apresenta as amostras de imagens obtidas após o pré-processamento.

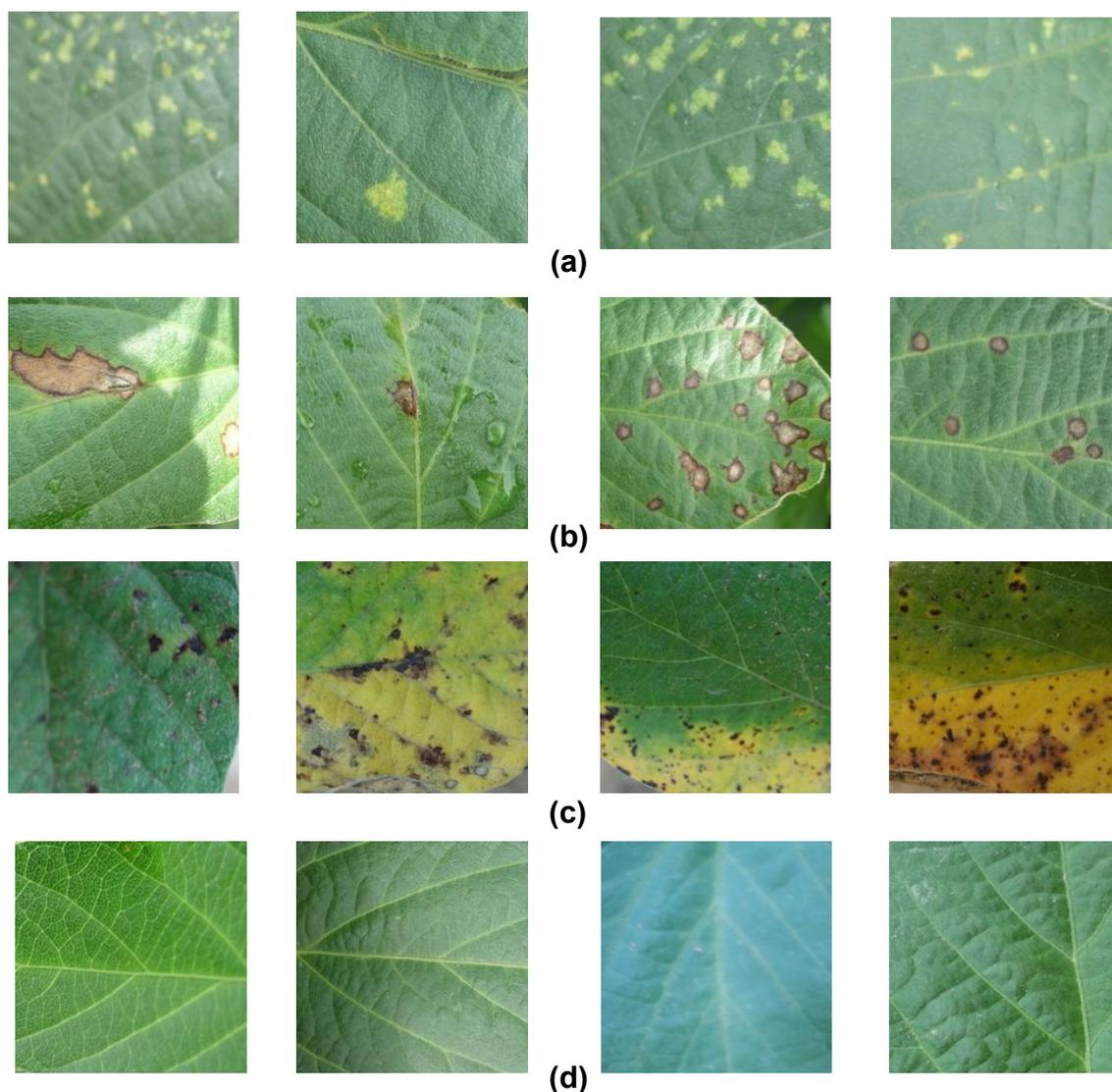


Figura 5.5: Conjunto de imagens das classes do problema de identificação de doenças foliares da cultura de soja. (a) Classe de folhas com a doença míldio, (b) Classe de folhas com a doença mancha olho de rã, (c) Classe de folhas com a doença da mancha parda, e (d) Classe de folhas saudáveis.

5.2 Resultados Experimentais

Com base nos bancos de imagens construídos na Seção 5.1 e técnicas propostas no Capítulo 4, foram realizados dois experimentos, sendo um para cada base de imagens de Soja. Os experimentos tem como objetivos principais avaliar o desempenho das técnicas propostas e realizar uma comparação com as técnicas de classificação que utilizam vocabulários visu-

ais descritas no Capítulo 3, bem como detectar combinações de técnicas e parâmetros que obtivessem os melhores desempenhos na solução do problema de identificação de doenças foliares na cultura de soja.

Em todos os experimentos, foi utilizado um computador disponível para o projeto na Fundação Universidade Federal de Mato Grosso do Sul, sede em Ponta Porã/MS, com a seguinte configuração: Processador Intel(R) Core(TM) i7-5820K 3.30 GHz, 32 GB de memória RAM, Placa mãe AsRock X99 Extreme4 LGA2011-V3, SSD M.2 Samsung 950 Pro 512GB, HD 2TB 7200RPM 64MB Cache e GPU NVIDIA Titan X disponibilizada pela NVIDIA.

Para atingir os objetivos propostos foram extraídos os descritores de pontos locais de cada imagem, dos respectivos bancos de imagens, através do SIFT [40] e do SIFT Denso [38]. Na etapa seguinte foi aplicado o algoritmo de agrupamento *K-Means* para construir os vocabulários com K palavras visuais, sendo cada palavra representada por um centroide. O valor de K foi variado de acordo com a quantidade de imagens do banco. Em seguida os descritores e os centroides foram submetidos ao BoVW, VLAD, VLAD-F, VLAD-PNC e ao VLAD-VNNC.

Para efeito de simplificação dos vetores VLAD, foi considerado o VLAD em sua terceira forma de normalização, chamada *componet-wise mass normalization*. Visto que, experimentos preliminares não apresentaram diferenças significativas entre as normalizações.

Para classificar as imagens foi utilizado o classificador *Linear Support Vector Machine - SVM* [39], por ser simples, intuitivo e fornecer resultados fáceis de serem interpretados. O tamanho dos conjuntos de treino e teste foram definidos de acordo com o tamanho do banco de imagens, sendo que, para o banco Soja Scanner foram utilizados conjuntos de 30, 50, 75, 100 e 150 imagens para treino; e, para o banco Plant Village foram utilizados conjuntos de 100, 300, 600 e 800 imagens para treino.

Para melhor avaliar o desempenho foram realizadas 10 repetições da classificação e foi calculada a média dos percentuais de acerto, o desvio padrão e a matriz de confusão. Os resultados obtidos são apresentados a seguir.

5.2.1 Resultados - Soja Scanner

O primeiro experimento tomou como base o banco de imagens descrito na Seção 5.1.1, composto por 1200 imagens de tamanho 200×200 pixels, em que 900 imagens correspondem a folhas doentes (ferrugem RB, ferrugem TAN e míldio) e 300 são folhas sadias. A base Soja Scanner, tem como principais características, imagens bem delimitadas e digitalizadas, o que permite uma qualidade superior nas imagens das amostras.

O objetivo desse experimento foi avaliar o desempenho das técnicas com relação ao tamanho do conjunto K , que corresponde a quantidade de palavras visuais que foram detectados para formar o vocabulário. Em cada uma das técnicas foram utilizados os valores de K variando entre: 8, 16, 32, 64, 128, 256 e 512.

A Tabela 5.1 apresenta os melhores resultados obtidos na tarefa do reconhecimento das doenças utilizando os descritores locais, SIFT e SIFT Denso, e todas as técnicas. Os resultados apresentados foram alcançados com o máximo de imagens para treino no SVM Linear, 150. Isso se deve ao fato que os classificadores SVM's baseiam-se em uma margem máxima entre as classes para classificar novos exemplos, e maximizando a margem, se melhora a generalização na fase de classificação.

Os resultados demonstram que as técnicas propostas VLAD-PNC com $P = 2$, e VLAD-VNNC com 80% dos descritores mais próximos, são superiores tanto ao BoVW, quanto ao VLAD, independente do tamanho dos conjuntos K , e do descritor de pontos locais utilizado. Todavia é importante ressaltar que, com exceção do VLAD-F e do BoVW, todas as técnicas apresentaram resultados acima de 90% na classificação.

O VLAD-F, que utiliza uma relação de todos os descritores com todas as palavras visuais, teve um desempenho satisfatório quando comparado ao BoVW utilizando o SIFT. No entanto, isso não se repetiu na comparação com o VLAD e o BoVW com SIFT Denso. A atribuição de todos os descritores para todos os centroides, proposta no VLAD-F, visa fornecer mais informações para o cálculo dos vetores VLAD. Porém, essa alternativa recaiu nos mesmos problemas encontrados em uma atribuição rígida, pois descritores que estavam muito distantes também foram rotulados aos centroides e as informações adicionadas afetaram negativamente no desempenho.

Tabela 5.1: Comparação do desempenho das técnicas, em relação ao tamanho dos Vocabulários no banco de imagens Soja Scanner, utilizando o SIFT e o SIFT Denso como descritores de pontos locais.

Técnicas X Tamanho de Vocabulários	8	16	32	64	128	256	512
SIFT							
BoVW	72.75 (±2.017)	79.47 (±2.368)	80.48 (±2.585)	82.83 (±1.269)	82.65 (±1.368)	84.22 (±1.451)	86.23 (±1.128)
VLAD	91.47 (±0.834)	92.57 (±0.551)	92.98 (±0.795)	93.27 (±0.759)	92.65 (±0.709)	92.07 (±0.896)	90.78 (±1.040)
VLAD-F	85.45 (±1.773)	86.65 (±1.857)	87.13 (±1.638)	87.15 (±1.586)	87.05 (±1.627)	87.28 (±1.718)	87.33 (±1.700)
VLAD PNC	92.73 (±0.615)	94.27 (±0.582)	95.03 (±0.514)	95.08 (±1.052)	94.70 (±0.804)	93.73 (±1.100)	92.70 (±1.099)
VLAD VNNC	93.98 (±0.590)	94.15 (±0.500)	94.77 (±0.738)	95.42 (±0.681)	96.12 (±0.698)	95.47 (±0.728)	95.03 (±0.936)
SIFT Denso							
BoVW	60.45 (±1.761)	66.22 (±5.513)	79.73 (±2.886)	88.20 (±0.728)	89.02 (±0.880)	91.10 (±0.568)	91.87 (±0.793)
VLAD	95.15 (±0.992)	96.10 (±0.876)	96.83 (±0.766)	97.12 (±0.754)	97.15 (±0.616)	97.03 (±0.628)	96.18 (±0.718)
VLAD-F	75.35 (±1.609)	76.07 (±1.612)	76.47 (±0.892)	76.08 (±1.958)	76.77 (±1.923)	77.12 (±1.503)	77.18 (±1.702)
VLAD PNC	95.10 (±0.840)	96.65 (±0.795)	96.60 (±0.956)	97.22 (±0.861)	97.40 (±0.672)	97.17 (±0.685)	96.95 (±0.809)
VLAD VNNC	94.60 (±0.649)	96.15 (±1.101)	97.53 (±0.939)	97.63 (±0.597)	97.43 (±0.717)	97.68 (±0.673)	97.42 (±0.712)

Ainda, com base nos resultados dos experimentos apresentados na Tabela 5.1, ao se analisar o descritor de pontos locais SIFT, o melhor valor foi obtido pelo VLAD-VNNC, com 96.12% de acerto na classificação utilizando um vocabulário de tamanho $K = 128$, o VLAD-PNC também obteve desempenho elevado com 95.08% de acerto e um vocabulário menor de tamanho $K = 64$. Porém o VLAD-F, mesmo com um vocabulário de 512 palavras teve um desempenho de 87.33%, sendo inferior ao VLAD original, que obteve 93.27% de acerto em um vocabulário de 64 palavras e pouco superior ao BoVW que atingiu 86.23% de acerto com vocabulário de $K = 512$. Além disso, o aumento da quantidade de palavras do vocabulário não gerou uma melhora significativa no desempenho, conforme se pode notar pela Tabela 5.1. Ao se aumentar o tamanho do vocabulário de 32 palavras para 128, no VLAD-VNNC, a melhoria no desempenho da classificação foi menor que 2%. E, em dado momento, tanto no VLAD, quanto no VLAD-PNC e VLAD-VNNC o desempenho da classificação veio a diminuir a medida que o tamanho dos vocabulários crescia, isso demonstra que o aumento da quantidade de palavras no vocabulário não gerou bons agrupamentos.

Dentre os resultados apresentados pelo descritor local SIFT Denso, o que obteve o melhor desempenho foi o VLAD-VNNC com taxa de classificação correta de 97.68% em um vocabulário com $K = 256$. No VLAD-PNC o melhor percentual de acerto foi 97.40% com $K = 128$, O VLAD também teve um desempenho convincente, sendo o melhor resultado 97.12% atingindo com $K = 128$. É necessário observar que todos os resultados obtidos com VLAD, VLAD-PNC e VLAD-VNNC se mantiveram acima de 94% de acerto para todos os valores de K avaliados.

O BoVW obteve o melhor desempenho com $K = 512$ e taxa de classificação correta igual a 91.87%, sendo que o valor das taxas de classificação evoluiu a medida que o tamanho de K aumentava para essa técnica. Já o VLAD-F não teve um desempenho satisfatório, a melhor taxa de classificação foi 77.18% obtido com $K = 512$, ficando bem abaixo de todas as demais técnicas, conforme pode ser observado na Tabela 5.1.

Ao se analisar o aumento no tamanho dos vocabulários junto ao SIFT Denso, se pode notar uma melhoria geral nos resultados a medida que o valor de K aumentava. Isso ocorre pois o SIFT Denso gera uma maior quantidade de pontos de interesse e desta forma o K -means consegue obter agrupamentos melhores e representar mais adequadamente o vocabulário.

O gráfico da Figura 5.6 apresenta as melhores taxas de classificação obtidas pelas técnicas, para a base Soja Scanner. Os resultados foram divididos de acordo com o descritor local utilizado (SIFT, SIFT Denso) e técnica de classificação de imagens utilizada.

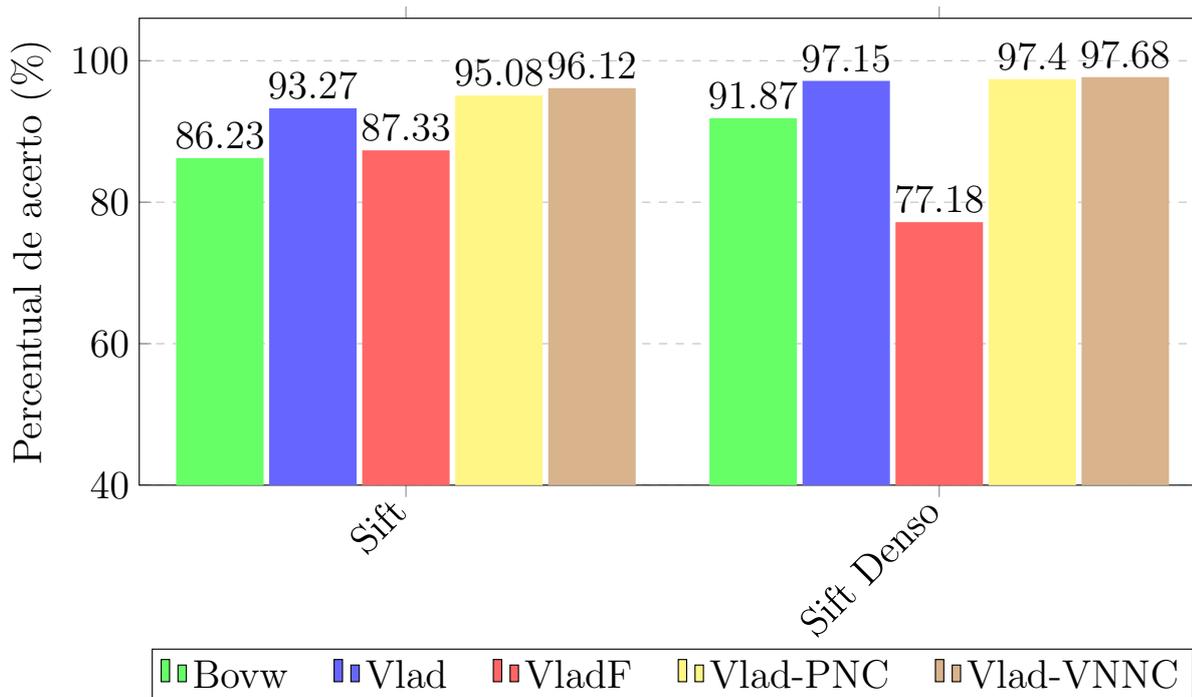


Figura 5.6: Taxa de classificação correta na base de imagens Soja Scanner, obtida pelas técnicas propostas utilizando os descritores SIFT e SIFT Denso.

A Figura 5.7 ilustra o desempenho das técnicas de acordo com as classes do problema de reconhecimento de doenças foliares da cultura de soja. Nas matrizes de confusão são considerados os valores presentes na diagonal principal, de forma que, quanto maior os valores presentes na diagonal, melhor é o desempenho.

No desempenho por classe, a classe míldio teve o melhor desempenho em todos os algoritmos analisados, por apresentar características mais nítidas, como manchas de coloração verde-claras até amareladas localizadas na face das folhas. Podemos observar que em quatro matrizes a taxa de acerto do míldio foi de 100%. As folhas saudáveis também foram detectadas com uma alta taxa de acerto na maioria dos resultados, com exceção do VLAD-F com SIFT

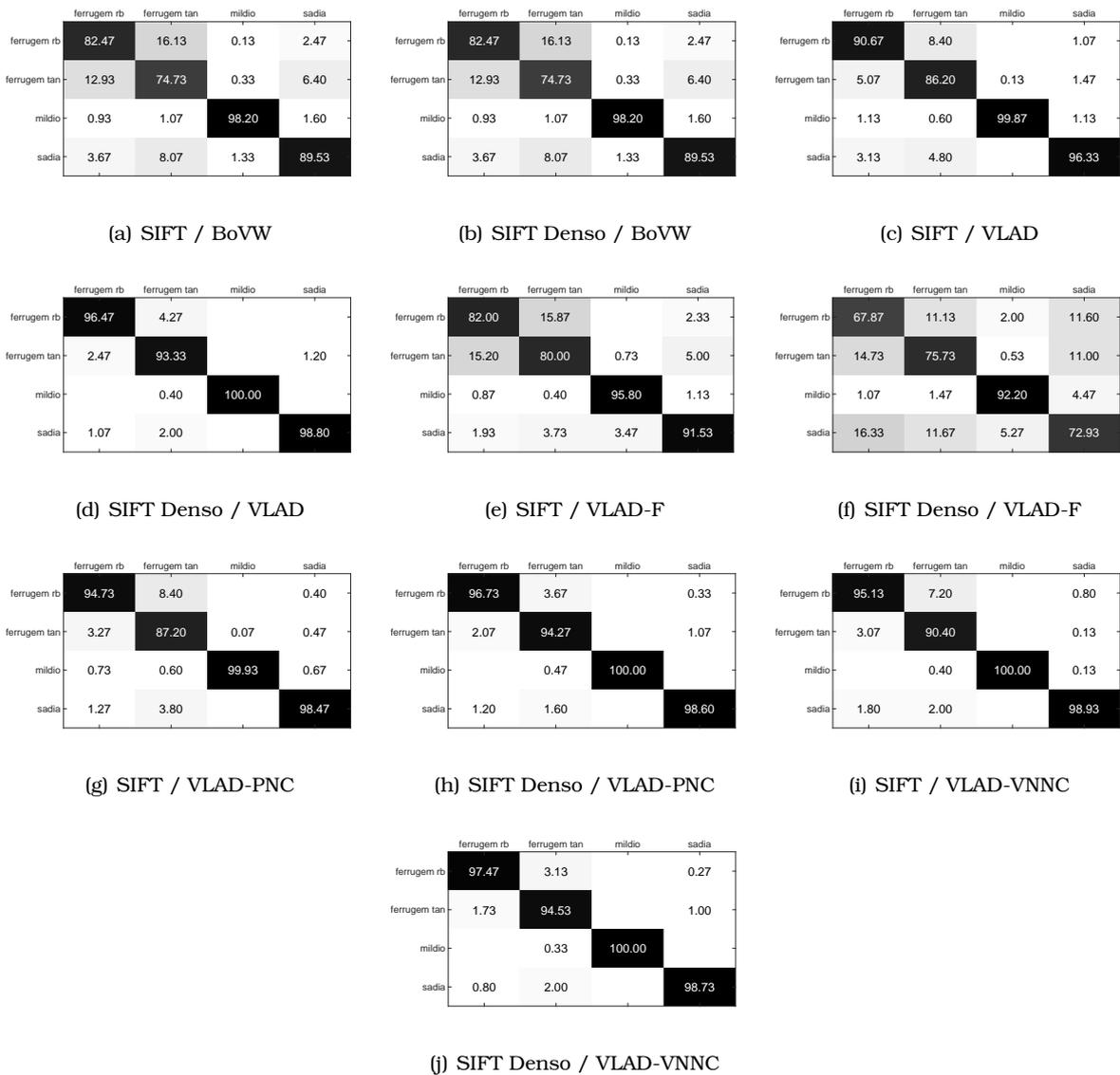


Figura 5.7: Matrizes de confusão que ilustram resultados obtidos na tarefa de classificação de doenças foliares junto a base de dados Soja Scanner 5.1.1.

Denso, que teve desempenho abaixo também na classificação das outras classes do problema. As imagens pertencentes as doenças ferrugem tan e ferrugem rb tiveram uma maior confusão entre si, devido a similaridade das suas imagens como pode ser observado na Figura 5.3.

5.2.2 Resultados - Plant Village

Nesta seção é descrito o segundo experimento, que tomou como base o banco de imagens descrito na Seção 5.1.2, composto por 7584 imagens de tamanho 199×201 pixels, em que 6439 imagens correspondem a folhas doentes (mildio, mancha olho de rã, mancha parda) e 1145 são folhas saudáveis. Entre as principais características desse banco estão a quantidade de amostras presentes, as diferentes regiões em que as imagens foram obtidas, as diferentes iluminações em que as imagens foram retiradas e a forma de obtenção que utilizou uma câmera "comum".

Nesse experimento, o principal propósito foi validar as técnicas, quanto a sua aplicação em situações reais, dadas as características das imagens. Para tal fim, foram utilizados os descritores locais SIFT e SIFT Denso, e foram construídos os vocabulários visuais com K

variado entre: 8, 16, 32, 64, 128, 256 e 512 tal como no experimento anterior; e, foi utilizado o classificador SVM Linear com 800 imagens para treino.

A Tabela 5.2 apresenta os melhores resultados obtidos pelas técnicas na tarefa de reconhecimento das doenças na base de dados Plant Village. Os resultados experimentais demonstram que as técnicas propostas tem aplicação em ambientes reais, uma vez que, obtiveram taxa de acerto satisfatórias, com resultados superiores a 90% em dadas combinações de parâmetros.

O melhor resultado obtido foi através da combinação do descritor local SIFT Denso, vocabulário construído com $K = 512$ no VLAD-VNNC que obteve uma taxa de acerto de 92.95%. Com relação aos descritores locais, o SIFT atingiu melhor resultado quando utilizado junto ao VLAD-VNNC e um vocabulário de tamanho 256 palavras, que obteve com uma taxa de acerto de 84.39%. O SIFT Denso alcançou seu melhor resultado com a combinação já mencionada, $K = 512$ e VLAD-VNNC.

Ao se comparar os resultados apresentas na Tabela 5.1 com os da Tabela 5.2, se pode notar que as características das imagens influenciaram diretamente no desempenho da classificação. Em termos absolutos os resultados na base Plant Village foram inferiores aos obtidos pelas técnicas sobre a base de imagens Soja Scanner, todavia, essa variação já era esperada dadas as diferenças entre as bases e o fato de que a base Plant Village ser mais próxima a condições reais encontradas no campo.

Com relação as técnicas, o VLAD-F não atingiu um desempenho satisfatório, tanto com o descritor local SIFT quanto com o SIFT Denso, os resultados foram 72.46% e 63.42% respectivamente. Isso ocorreu devido a proposta do VLAD-F de relacionar todos os descritores com todas as palavras visuais.

O BoVW teve um desempenho relativamente baixo com vocabulários pequenos entre 8, 16 e 32 palavras, porém a medida que o tamanho K dos vocabulários crescia, o desempenho também evoluiu. Os resultados também foram influenciados pelos descritores, de forma que, os melhores percentuais foram obtidos utilizando o descritor de ponto SIFT Denso conforme podemos observar na Tabela. O VLAD manteve seu desempenho suficiente para a base, e também demonstrou evolução no percentual de acerto a medida que o tamanho dos vocabulários crescia.

O VLAD-PNC e o VLAD-VNNC apresentaram os melhores desempenhos para a base de imagens Plant Village, com o VLAD-VNNC um pouco superior. Ambas as abordagens propõe a atribuir um descritor a mais do que uma palavra visual com base no calculo do peso que o descritor exerce sobre o centroide, fornecendo um conjunto maior de informações para a construção dos vetores.

Com o descritor local SIFT o pior resultado do VLAD-PNC foi 78.55% com $K = 8$, o melhor desempenho foi obtido com $K = 512$ sendo de 84.02% e média foi em torno de 81.94%. No SIFT Denso, o pior resultado do VLAD-PNC foi 76.39% com $K = 8$, e o melhor desempenho foi 91.98% com vocabulário de tamanho $K = 512$; o desempenho médio foi de 84.90%.

No VLAD-VNNC, utilizando o descritor local SIFT a menor taxa de acerto foi 77.85% com $K = 8$; a melhor taxa foi 84.39% com o vocabulário $K = 256$; o desempenho médio foi de 82.28%. Utilizando descritor local SIFT Denso, o menor desempenho foi 75.39% com $K = 8$ e o melhor foi 92.95% com $K = 512$, a média ficou em 86.22%.

Ao se analisar o crescimento dos vocabulários com relação ao desempenho para todas as técnicas avaliadas, se pode notar que no geral o aumento do tamanho dos vocabulários levou ao crescimento do desempenho, exceto pelo VLAD-F. No entanto, esse crescimento não se deu de forma proporcional em todas as técnicas. A maior taxa de aumento no desempenho ocorreu no BoVW com o descritor SIFT Denso, em que a diferença entre o desempenho dos vocabulários $K = 512$ para $K = 8$ foi de 39.71%.

Tabela 5.2: Comparação do desempenho das técnicas, em relação ao tamanho dos Vocabulários no banco de imagens Plant Village, utilizando o SIFT e o SIFT Denso como descritores de pontos locais.

Técnicas X Tamanho de Vocabulários	8	16	32	64	128	256	512
SIFT							
BoVW	55.89 (±4.878)	64.73 (±4.886)	71.34 (±2.803)	73.40 (±1.928)	74.24 (±2.667)	75.64 (±0.816)	76.47 (±0.537)
VLAD	75.77 (±0.718)	77.88 (±0.515)	78.34 (±0.722)	79.87 (±0.542)	80.67 (±0.353)	82.03 (±0.516)	81.71 (±0.557)
VLAD-F	70.76 (±1.766)	70.76 (±0.976)	71.86 (±0.857)	72.22 (±0.628)	72.01 (±0.694)	72.46 (±0.895)	72.28 (±0.884)
VLAD PNC	78.55 (±0.272)	80.28 (±0.505)	81.78 (±0.365)	82.30 (±0.524)	83.07 (±0.587)	83.75 (±0.421)	84.02 (±0.547)
VLAD VNNC	77.85 (±0.525)	80.43 (±0.601)	81.93 (±0.632)	83.16 (±0.699)	83.96 (±0.690)	84.39 (±0.425)	84.25 (±0.429)
SIFT Denso							
BoVW	47.31 (±8.931)	50.42 (±12.598)	60.38 (±6.695)	78.27 (±4.064)	81.53 (±1.343)	84.97 (±0.812)	87.02 (±0.377)
VLAD	75.43 (±0.400)	79.04 (±0.673)	85.42 (±0.435)	88.67 (±0.372)	90.49 (±0.427)	90.86 (±0.333)	90.55 (±0.342)
VLAD-F	58.45 (±6.906)	61.07 (±6.255)	63.42 (±3.455)	60.30 (±8.164)	60.65 (±3.851)	62.51 (±5.917)	60.34 (±4.984)
VLAD PNC	76.39 (±0.604)	79.62 (±0.559)	84.91 (±0.533)	88.76 (±0.442)	90.76 (±0.406)	91.84 (±0.393)	91.98 (±0.398)
VLAD VNNC	75.39 (±0.808)	79.05 (±0.380)	84.69 (±0.390)	88.40 (±0.436)	90.77 (±0.393)	92.29 (±0.338)	92.95 (±0.368)

Porém, quando avaliados os desempenhos dos vocabulários de tamanho $K = 256$ para $K = 512$ a melhoria dos percentuais de acerto foi em geral baixa. Isso indica que aumento dos vocabulários pode levar a construção de agrupamentos não tão representativos. Conforme podemos observar na Tabela 5.2 o desempenho do VLAD-PNC com 256 palavras foi de 91.84% e com 512 palavras foi de 91.98%, gerando uma melhoria de 0.14%, que é muito baixa se considerar a diferença envolvida entre os vocabulários.

O gráfico da Figura 5.8 apresenta as melhores taxa de classificação obtidas pelas técnicas para a base Plant Village. Os resultados foram divididos de acordo com o descritor local utilizado (SIFT, SIFT Denso) e técnica utilizada.

A Figura 5.9 ilustra as matrizes de confusão do desempenho das técnicas de acordo com as classes do problema. Como esperado o VLAD-F apresentou uma maior confusão na classificação das imagens, se pode observar que os valores presentes na diagonal principal, tanto do SIFT quanto do SIFT Denso, são os menores dentre todas as matrizes.

O VLAD-VNNC e o VLAD-PNC, tal como previsto, tiveram as menores taxa de confusão visto que os valores presentes em suas diagonais principais são altos. Também se pode notar uma vantagem com relação ao descritor de pontos SIFT Denso, conforme podemos observar na Figura 5.9 (g), (h), (i) e (j). O BoVW e o VLAD tiveram desempenhos satisfatórios, e apresentaram matrizes de confusão bem definidas em suas diagonais principais.

No desempenho por classe, a classe das imagens da Mancha parda apresentaram os melhores resultados para a maioria das técnicas analisadas, isso ocorreu devido a quantidade de imagens que foram analisadas e as características das imagens, visto que as amostras da Mancha parda totalizavam 3565 imagens, marcadas por manchas amarelas fortes com pontos

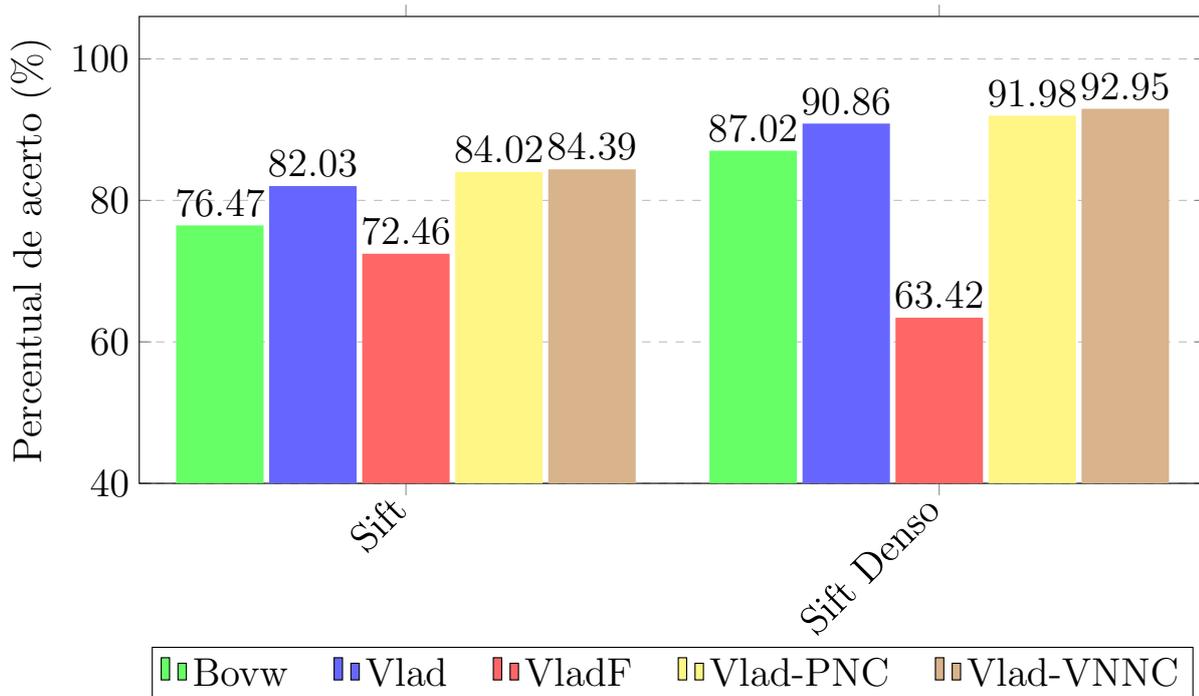


Figura 5.8: Taxa de classificação correta na base de imagens Plant Village, obtida pelas técnicas propostas utilizando os descritores SIFT e SIFT Denso.

escuras. O melhor resultado foi obtido pelo descritor local SIFT Denso com o VLAD-VNNC.

A segunda classe com melhor desempenho foram as imagens do Míldio, que apesar de contar com o conjunto menor de amostras, 851 imagens, apresenta imagens com características bem definidas por manchas amareladas espalhadas na face das folhas, as imagens dessa classe apresentaram um pouco de confusão com as amostras de imagens da classe Olho de rã e das imagens Sadias. A terceira classe melhor identificada foi a das amostras Sadias, que é composta por 1145 imagens sem traços de doenças. As imagens dessa classe apresentam folhas com coloração verde, sem ranhuras ou manchas.

As maiores confusões ocorrem nas imagens da classe Olho de rã, que conta com 2023 amostras marcadas por manchas marrons na face das folhas, essa classe apresentou uma maior confusão com as imagens da classe Sadia e da classe Míldio conforme pode ser observado na Figura 5.9.

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	80.59	9.24	2.43	4.43
Olho de rã	10.20	60.24	9.08	20.84
Mancha parda	4.12	9.57	84.35	4.52
Sadia	5.10	20.96	4.13	70.20

(a) SIFT / BoVW k

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	85.69	9.75	0.53	3.25
Olho de rã	10.00	70.26	3.59	15.45
Mancha Parda	1.18	3.52	95.20	0.32
Sadia	3.14	16.48	0.68	80.99

(b) SIFT Denso / BoVW

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	87.45	6.74	1.63	1.71
Olho de rã	6.27	66.13	5.55	11.74
Mancha parda	2.16	5.05	88.64	1.94
Sadia	4.12	22.08	4.18	84.61

(c) SIFT / VLAD

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	96.86	4.01	0.69	1.10
Olho de rã	1.76	80.53	2.48	5.33
Mancha parda		1.31	94.58	0.64
Sadia	1.37	14.15	2.25	92.93

(d) SIFT Denso/ VLAD

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	74.12	15.65	4.46	6.12
Olho de rã	16.47	52.07	8.17	19.94
Mancha parda	5.88	9.99	81.90	5.16
Sadia	3.53	22.29	5.47	68.78

(e) SIFT / VLAD-F

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	49.22	19.90	8.34	11.39
Olho de rã	20.00	36.91	10.69	23.59
Mancha parda	17.45	15.87	74.72	7.65
Sadia	13.33	27.32	6.25	57.36

(f) SIFT Denso/ VLAD-F

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	90.78	5.17	1.29	1.65
Olho de rã	5.29	67.96	4.47	8.72
Mancha parda	1.18	4.61	90.57	2.17
Sadia	2.75	22.26	3.67	87.45

(g) SIFT / VLAD-PNC

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	95.88	3.07	0.56	0.67
Olho de rã	2.35	83.61	1.91	5.88
Mancha parda		1.09	95.50	0.49
Sadia	1.76	12.24	2.04	92.96

(h) SIFT Denso / VLAD-PNC

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	89.80	4.98	1.39	1.86
Olho de rã	5.10	69.70	4.44	9.22
Mancha parda	1.18	4.29	90.50	2.20
Sadia	3.92	21.03	3.67	86.72

(i) SIFT / VLAD-VNNC

	Mildio	Olho de rã	Mancha parda	Sadia
Mildio	96.08	2.94	0.40	0.99
Olho de rã	2.55	84.99	1.80	5.04
Mancha parda		1.09	96.33	0.35
Sadia	1.37	10.99	1.47	93.62

(j) SIFT Denso/ VLAD-VNNC

Figura 5.9: Matrizes de confusão que ilustram resultados obtidos na tarefa de classificação de doenças foliares junto a base de dados Plant Village 5.1.2.

Conclusão

Neste trabalho foi proposto um conjunto de melhorias nos algoritmos de classificação de imagens que permitiram a sua aplicação no reconhecimento de doenças foliares da cultura de soja em imagens de lavouras. Foram propostas abordagens baseadas no Histograma de Palavra Visuais (BoVW) e nos Vetores Localmente Agregados (VLAD), que possibilitaram a inclusão de informações locais adicionais aos descritores das imagens. Foi aplicado o cálculo de pesos para cada descritor com relação as palavras visuais e os descritores foram atribuídos aos centroides em que exercessem maior peso.

As técnicas foram avaliadas sobre dois bancos de imagens compostos por amostras de doenças foliares da soja. Sendo que, o primeiro banco é composto por 1200 imagens obtidas com o auxílio de um *scanner*, e o segundo é composto por 7584 imagens obtidas através de uma câmera em campos de experimentos. Foram utilizados os descritores de pontos locais SIFT e SIFT Denso, para obter e descrever os pontos de interesse nas imagens, em seguida foi utilizado o algoritmo de agrupamento *K-means* para a construção dos vocabulários visuais. Após esta etapa foram realizados os cálculos de peso dos descritores e as atribuições de acordo com a técnica aplicada.

Os resultados experimentais demonstraram que a inclusão de informações espaciais locais aos descritores melhora o desempenho na tarefa de reconhecimento e classificação de doenças foliares da cultura de soja. Dentre as técnicas propostas VLAD-PNC e VLAD-VNNC obtiveram os melhores desempenhos para ambas as bases e com ambos os descritores, superando as técnicas propostas na literatura BoVW e VLAD.

Os resultados também comprovaram que o descritor de pontos locais SIFT Denso superou de maneira geral o SIFT. A exceção foi observada no VLAD-F, que teve resultados inferiores na maioria dos testes. Isso ocorreu pois o SIFT Denso extrai um conjunto maior de pontos de interesse para serem descritos localmente, desta forma o *K-means* com mais pontos consegue representar mais adequadamente o vocabulário. Para o primeiro banco de imagens, o SIFT Denso extraiu 2209 pontos por imagem do banco, enquanto que o SIFT extraiu em média 354. Para a segunda base, o SIFT Denso extraiu, novamente, 2209 pontos por imagens, e o SIFT obteve em média 191 pontos.

Na primeira base de imagens o VLAD-VNNC utilizando o SIFT Denso e um vocabulário de 256 palavras atingiu 97.68% de acerto na classificação. Resultado pouco superior ao obtido

pelo VLAD-PNC, também com o descritor de pontos SIFT Denso e um vocabulário composto por 128 palavras visuais que obteve 97.40% de taxa de acerto na classificação. Na segunda base de imagens, Plant Village, o VLAD-VNNC obteve uma taxa de acerto de 92.95%, utilizando o descritor local SIFT Denso com um vocabulário de 512 palavras. Por outro lado, o VLAD-PNC alcançou 91.98%, com o descritor SIFT Denso e um vocabulário de 512 palavras.

Com base nos resultados obtidos na segunda base de imagens foi possível concluir que as técnicas tem aplicação para problemas reais, visto que está base simula mais fielmente as condições encontradas nos cultivos, que dificultam o processo de reconhecimento das doenças.

6.1 *Trabalhos Futuros*

Como proposta para trabalhos futuros, podemos elencar:

- Avaliar as técnicas propostas com relação a utilização de outros descritores de pontos locais, como por exemplo: SURF, PHOW, HOG e o LBP;
- Avaliar as técnicas propostas com relação a utilização de outros algoritmos de agrupamentos, como: GMM, Florestas de agrupamentos randômicos, Florestas de caminhos ótimos. Que tem demonstrados resultados promissores na substituição do *K-means*, pelo fato de não precisarem conhecer previamente o tamanho do vocabulário (K);
- Desenvolver um módulo que permita identificar e analisar a severidade das doenças foliares, possibilitando a padronização do processo que hoje, é realizado visualmente por especialistas da área estando sujeito a subjetividade;
- Aplicar e avaliar o método proposto para outras culturas como: laranja, cana-de-açúcar, milho, algodão. Visto que estas culturas estão muito presentes no ambiente agrícola de nosso país e demandam cuidados com relação a perdas provocadas por doenças.

Referências Bibliográficas

- [1] Plano agrícola e pecuário 2014/2015. Technical report, Ministério da Agricultura, Pecuária e Abastecimento, Brasília/DF, 2015. <http://www.agricultura.gov.br/>. Citado na página 1.
- [2] Luis Claudio Sugi Afonso. Composição de dicionários visuais utilizando agrupamento de dados por florestas de caminhos ótimos. 2012. Citado na página 7.
- [3] Relja Arandjelovic and Andrew Zisserman. All about vlad. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1578–1585, 2013. Citado na página 32.
- [4] Dana H Ballard and Christopher M Brown. Computer vision, article, 4 pages prentice-hall. *Englewood Cliffs, New Jersey, believed to be published more than one year prior to the filing date of the present application*, 1982. Citado na página 16.
- [5] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In *Computer vision—ECCV 2006*, pages 404–417. Springer, 2006. Citado nas páginas 3, 7, e 26.
- [6] Emidio Rizzo Bonato and Ana Lidia Variani Bonato. A soja no brasil: história e estatística. *Embrapa Soja-Docmentos (INFOTECA-E)*, 1987. Citado na página 9.
- [7] Marcelo Rafael Borth, Julio Cezar Iacia, Hemerson Pistori, and Clandio Favarini Ruviaro. A visão computacional no agronegócio: Aplicações e direcionamentos. *7º Encontro Científico de Administração, Economia e Contabilidade (ECAECO)*, 2014. Citado nas páginas 3 e 17.
- [8] Anna Bosch, Andrew Zisserman, and Xavier Muoz. Image classification using random forests and ferns. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007. Citado na página 3.
- [9] Matthew Brown and David G Lowe. Invariant features from interest point groups. In *BMVC*, number s 1, 2002. Citado na página 20.
- [10] C.R. Casela. *Doenças da soja*. EMBRAPA-UEPAE de Pelotas. Circular técnica. EMBRAPA-UEPAE de Pelotas, 1979. Citado na página 2.
- [11] Yixin Chen and James Z Wang. Image categorization by learning and reasoning with regions. *The Journal of Machine Learning Research*, 5:913–939, 2004. Citado na página 6.

- [12] CONAB. Acompanhamento da safra brasileira de grãos. Technical report, Companhia Nacional de Abastecimento., Brasília/DF, 2015. <http://www.conab.gov.br>. Citado nas páginas xi, 1, e 2.
- [13] José Alfredo Ferreira Costa and Márcio Luiz de Andrade Netto. Segmentação de mapas auto-organizáveis com espaço de saída 3-d. *Sba: Controle & Automação Sociedade Brasileira de Automatica*, 18(2):150–162, 2007. Citado na página 16.
- [14] Paulo E Cruvinel and Décio Karam. Método de reconhecimento de padrões de plantas invasoras de folhas largas e estreitas da cultura do milho (zea mays l.) baseado. Citado na página 17.
- [15] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2. Prague, 2004. Citado nas páginas 3, 6, 8, e 25.
- [16] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005. Citado na página 3.
- [17] Antonio Carlos de Nazaré Júnior, David Menotti, José Maria Ribeiro Neves, and Tuneo Sedyama. Detecção automática da área foliar danificada da soja através de imagens digitais. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2009. Citado nas páginas 3 e 17.
- [18] JM Dunleavy et al. Yield reduction in soyabeans caused by downy mildew. *Plant disease*, 71(12):1112–1114, 1987. Citado na página 13.
- [19] C F Estrada, A Jepsen, and D Fleet. Local features tutorial 2. 2004. Citado nas páginas xi, xii, 23, e 24.
- [20] JRB Farias, AL Nepomuceno, and N Neumaier. Ecofisiologia da soja. *Embrapa Soja. Circular técnica*, 2007. Citado na página 10.
- [21] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 524–531. IEEE, 2005. Citado na página 6.
- [22] Felipe Cesar Stanzani Fonseca and Walber Antônio Ramos Beltrame. Aplicações práticas dos algoritmos de clusterização k-means e bisecting k-means. *Universidade Federal do Espírito Santo (UFES). Vitória-ES*. Citado nas páginas 3 e 27.
- [23] PK Goel, SO Prasher, RM Patel, A DiTommaso, DL Smith, et al. Airborne optical remote sensing for detection of weed infestation in site specific management of field crops. In *2000 ASAE Annual International Meeting, Milwaukee, Wisconsin, USA, 9-12 July 2000*, pages 1–13. American Society of Agricultural Engineers, 2000. Citado nas páginas 3 e 17.
- [24] Giancarlo Luis Gómez Gonzáles. *Aplicação da Técnica SIFT para Determinação de Campos de Deformações de Materiais usando Visão Computacional*. PhD thesis, PUC-Rio, 2010. Citado nas páginas xii, 19, e 25.
- [25] José Fernando Jurca Grigoll. Manejo de doenças na cultura da soja. Technical report, Fundação MS., Campo Grande/MS, 2014/2015. <http://www.fundacaoms.org>.

br/tecnologia-producao-soja-2014-2015. Citado nas páginas xi, 10, 11, 12, 13, e 14.

- [26] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988. Citado na página 21.
- [27] Ademir Assis Henning, Álvaro Manuel Rodrigues Almeida, Cláudia Vieira Godoy, Claudine Dinali Santos Seixas, José Tadashi Yorinori, Leila Maria Costamilan, Léo Pires Ferreira, Maurício Conrado Meyer, Rafael Moreira Soares, and Waldir Pereira Dias. *Manual de identificação de doenças de soja*. Embrapa Soja, 2005. Citado nas páginas 10, 11, e 13.
- [28] Huy Tho Ho and Rama Chellappa. Automatic head pose estimation using randomly projected dense sift descriptors. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 153–156. IEEE, 2012. Citado na página 25.
- [29] Antônio Houaiss. *da língua portuguesa*. Rio de Janeiro, Objetiva, 2001. Citado na página 2.
- [30] David Hughes, Marcel Salathé, et al. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*, 2015. Citado na página 43.
- [31] Herve Jegou, Matthijs Douze, and Cordelia Schmid. Product quantization for nearest neighbor search. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(1):117–128, 2011. Citado na página 30.
- [32] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3304–3311. IEEE, 2010. Citado nas páginas 4, 6, 7, 16, 30, 31, e 33.
- [33] FC Juliatti, AC Polizel, and FC Juliatti. *Manejo integrado de doenças na cultura da soja. Uberlândia: UFU*, 2004. Citado na página 11.
- [34] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 506–513. IEEE Computer Society, 2004. Citado na página 25.
- [35] Patricia Fabretti Kreycki and José Otávio M. Menten. *Limitadoras de produtividade*. Technical report, Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo. Citado na página 2.
- [36] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178. IEEE, 2006. Citado na página 6.
- [37] Tony Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270, 1994. Citado na página 18.
- [38] Ce Liu, Jenny Yuen, and Antonio Torralba. Sift flow: Dense correspondence across scenes and its applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):978–994, 2011. Citado nas páginas 3, 8, 18, 25, 41, e 46.

- [39] Ana Carolina Lorena and André CPLF de Carvalho. Uma introdução às support vector machines. *Revista de Informática Teórica e Aplicada*, 14(2):43–67, 2007. Citado nas páginas 5 e 46.
- [40] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. Citado nas páginas xi, 3, 7, 8, 16, 18, 19, 20, 21, 22, 23, 26, 41, e 46.
- [41] JGR Maia. Detecção e reconhecimento de objetos usando descritores locais. *PhD in Bioinformatics, Universidade Federal do Ceara, Biomathematics Group, R. Qta. Grande*, 6:2780–156, 2010. Citado na página 16.
- [42] Frank Moosmann, Bill Triggs, and Frederic Jurie. Fast discriminative visual codebooks using randomized clustering forests. In *Twentieth Annual Conference on Neural Information Processing Systems (NIPS'06)*, pages 985–992. MIT Press, 2007. Citado na página 6.
- [43] Claudio Mário MUNDSTOCK and André Luis THOMAS. Soja: fatores que afetam o crescimento e o rendimento de grãos. *Porto Alegre: Universidade Federal do Rio Grande do Sul*, 2005. Citado nas páginas 9 e 10.
- [44] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002. Citado na página 3.
- [45] Jonatan Patrick Margarido Oruê, Wesley Eiji Sanches Kanashiro, Rillian Diello Lucas Pires, Wesley Nunes Gonçalves, Bruno Brandoli Machado, and Mauro dos Santos de Aruda. Identification of foliar soybean diseases using local descriptors. Citado na página 25.
- [46] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007. Citado na página 6.
- [47] Florent Perronnin, Christopher Dance, Gabriela Csurka, and Marco Bressan. Adapted vocabularies for generic visual categorization. In *Computer Vision–ECCV 2006*, pages 464–475. Springer, 2006. Citado na página 6.
- [48] Marion Pfeifer, Veronique Lefebvre, Toby A Gardner, Victor Arroyo-Rodriguez, Lander Baeten, Cristina Banks-Leite, Jos Barlow, Matthew G Betts, Joerg Brunet, Alexis Cerezo, et al. Biofrag—a new database for analyzing biodiversity responses to forest fragmentation. *Ecology and evolution*, 4(9):1524–1537, 2014. Citado na página 15.
- [49] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. Citado nas páginas 3 e 29.
- [50] EC Picinini and JM Fernandes. *Doenças de soja: Diagnose, epidemiologia e controle*. Embrapa Trigo, 2000. Citado na página 13.
- [51] Raquel Ramos Pinho, João Manuel Ribeiro Silva Tavares, and Miguel Fernando Paiva Velhote Correia. Introdução à análise de movimento usando visão computacional. 2004. Citado na página 16.

- [52] João Leonardo Fernandes Pires, Gilberto Rocca da Cunha, Aldemir Pasinato, Solange França, and Lisandro Rambo. *Discutindo agricultura de precisão-aspectos gerais*. Embrapa Trigo, 2004. Citado na página 2.
- [53] Rillian Diello Lucas Pires, Diogo Nunes Gonçalves, Jonatan Patrick Margarido Oruê, Wesley Eiji Sanches Kanashiro, Jose F Rodrigues, Bruno Brandoli Machado, and Wesley Nunes Gonçalves. Local descriptors for soybean disease recognition. *Computers and Electronics in Agriculture*, 125:48–55, 2016. Citado nas páginas xii, 41, e 42.
- [54] Simon JD Prince. *Computer vision: models, learning, and inference*. Cambridge University Press, 2012. Citado nas páginas 3 e 17.
- [55] Ana Paula de Oliveira Santos et al. Desenvolvimento de descritores de imagens para reconhecimento de padrões de plantas invasoras (folhas largas e folhas estreitas). 2009. Citado nas páginas 3 e 17.
- [56] Gercina Gonçalves da Silva, Kleber Padovani de Souza, Ariadne Barbosa Gonçalves, Hermerson Pistori, and Jenifer Ferreira Gonzaga. Veículos aéreos não tripulados com visão computacional na agricultura: Aplicações, desafios e perspectivas. *Anais do Encontro Científico de Administração, Economia e Contabilidade*, 1(1), 2015. Citado na página 16.
- [57] Olavo Corrêa da Silva. Danos causados pelo míldio da soja e uso de fosfitos e acibenzolar-s-methyl no manejo das doenças da cultura. 2013. Citado na página 13.
- [58] IMR Souza and GS Silva. Fungos associados a plantas daninhas na ilha de são luiz, maranhão. *Summa Phytopathologica*, 27(2), 2001. Citado na página 13.
- [59] Junior Silva Souza. Identificação de viabilidade de leveduras com corante vital utilizando histogramas de palavras visuais em imagens coloridas. 2015. Citado nas páginas xii, 3, 17, e 27.
- [60] Charbel Szymanski et al. Desenvolvimento de técnicas de processamento digital de imagens para inspeção de placas de circuito impresso produzidas em pequenas séries. 2014. Citado nas páginas xi e 23.
- [61] DE PRODUÇÃO DE SOJA REGIÃO TECNOLOGIAS. Central do brasil, 2003. *Londrina: Embrapa Soja*. Citado na página 9.
- [62] Leonardo Torok and Aura Conci. Método de otsu. *Instituto de Computação–Universidade Federal Fluminense, Niterói–RJ–Brasil, notas de aula*, 2014. Citado na página 17.
- [63] Jan C van Gemert, Cor J Veenman, Arnold WM Smeulders, and J-M Geusebroek. Visual word ambiguity. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(7):1271–1283, 2010. Citado nas páginas 4, 6, e 33.
- [64] A Vargas, Carlos ABCW Madsen, and Silvia SC Botelho. Navision-sistema de visão subaquática para navegação e montagem de mosaicos em auvs. In *Seminário e Workshop em Engenharia Oceânica*, 2004. Citado na página 16.
- [65] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong. Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3360–3367. IEEE, 2010. Citado na página 7.

- [66] XB Yang, AT Tschanz, WM Dowler, and TC Wang. Development of yield loss models in relation to reductions of components of soybean infected with phakopsora pachyrhizi. *Phytopathology*, 81(11):1420–1426, 1991. Citado na página 11.
- [67] JT Yorinori. Riscos de surgimento de novas doenças na cultura da soja. In *Congresso de Tecnologia e Competitividade da Soja no Mercado Global*, pages 165–169. Fundação Mato Grosso Cuiabá, 2000. Citado na página 2.
- [68] Kai Yu, Tong Zhang, and Yihong Gong. Nonlinear learning using local coordinate coding. In *Advances in neural information processing systems*, pages 2223–2231, 2009. Citado na página 7.
- [69] Jianguo Zhang, Marcin Marszałek, Svetlana Lazebnik, and Cordelia Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision*, 73(2):213–238, 2007. Citado na página 6.