UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS

Daniel Kenji Matsumoto Viana

Identificação de plantas a partir da análise de suas folhas utilizando o algoritmo SIFT e momentos de cor

São Carlos

2017

Daniel Kenji Matsumoto Viana

Identificação de plantas a partir da análise de suas folhas utilizando o algoritmo SIFT e momentos de cor

Monografia apresentada ao Curso de Engenharia Elétrica com Ênfase em Eletrônica, da Escola de Engenharia de São Carlos da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Engenheiro Eletricista.

Orientador: Prof. Dr. Marcelo Andrade da Costa Vieira

São Carlos

2017

Viana, Daniel Kenji Matsumoto
V614i Identificação de plantas a partir da análise de suas folhas utilizando o algoritmo SIFT e momentos de cor / Daniel Kenji Matsumoto Viana; orientador Marcelo Andrade da Costa Vieira. São Carlos, 2017.

Monografia (Graduação em Engenharia Elétrica com ênfase em Eletrônica) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2017.

1. identificação de plantas. 2. sift. 3. momentos de cor. I. Título.

FOLHA DE APROVAÇÃO

Nome: Daniel Kenji Matsumoto Viana

Título: "Identificação de plantas através da análise de suas folhas pelo algoritmo SIFT e por momentos de cor"

Trabalho de Conclusão de Curso defendido e aprovado em 01 / 12 / 2017 ,

com NOTA 10,0 (DEZ , ZERO), **pela Comissão Julgadora:**

Prof. Dr. Marcelo Andrade da Costa Vieira - Orientador - SEL/EESC/USP

Prof. Associado Evandro Luis Linhari Rodrigues - SEL/EESC/USP

Mestre Helder Cesar Rodrigues de Oliveira - Doutorando - SEL/EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Prof. Associado Rogério Andrade Flauzino

Este trabalho é dedicado aos meus pais, A. & S., e a todas as pessoas que me ajudaram e apoiaram durante a minha graduação.

AGRADECIMENTOS

Agradeço aos meus pais, A. e S., às minhas tias, S. e S., e à minha família pelo apoio durante a minha graduação.

Ao meu orientador Prof. Marcelo Andrade da Costa Vieira pela paciência e boa-vontade na elaboração da monografia.

Aos amigos e pessoas que tive oportunidade e prazer de conhecer em São Carlos.

Abre bem os olhos e vê.

Júlio Verne, Miguel Strogoff

Georges Perec, A vida: modo de usar

RESUMO

VIANA, D. K. M. **Identificação de plantas a partir da análise de suas folhas utilizando o algoritmo SIFT e momentos de cor**. 2017. 77p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2017.

A identificação de plantas é um problema antigo de agrônomos, agricultores, biólogos e pesquisadores, além de diletantes e do público em geral. É um processo laborioso que é feito manualmente, requerendo grande quantidade de tempo e experiência. Dessa maneira, não é um processo eficiente, perdendo-se muito tempo, especialmente quando se deve fazer a identificação de um conjunto grande de amostras. O presente trabalho tem como objetivo desenvolver um método baseado em visão computacional para realizar automaticamente a classificação de plantas, a partir da análise da imagem da folha da planta que se deseja identificar. A escolha da análise das folhas para a identificação das plantas foi feita pela maior facilidade de manuseio e maior disponibilidade delas, em comparação com outros métodos mais convencionais (análise das flores ou frutos). As características das folhas são obtidas pelos descritores do algoritmo SIFT e pela análise dos momentos de cor. Em seguida, é feito o agrupamento dos descritores através do método *k-means*. A análise da frequência de ocorrência desses grupos nas imagens segue um modelo *bag-of-words* de representação, que serve de entrada para uma máquina de vetores de suporte (SVM). A máquina, treinada com uma base de dados, realiza a identificação da espécie da planta. Por fim, o uso do método desenvolvido permitiu a comparação do desempenho do descritor SIFT puro e do descritor SIFT junto com os momentos de cor, para agrupamentos com diferentes números de centroides: no primeiro caso, a maior acurácia foi de 77,71%, utilizando 500 centroides; no segundo, a maior acurácia foi de 80,21%, utilizando 300 centroides. A adição dos momentos de cor nos casos estudados proporcionou um aumento médio da acurácia de 1,75%, com desvio padrão de 4,25% em relação aos casos somente com o descritor SIFT.

Palavras-chave: Visão computacional. Identificação de plantas. Análise de folhas. SIFT. Momentos de cor. Bag-of-words. SVM.

ABSTRACT

VIANA, D. K. M. **Plant identification by applying the SIFT algorithm and color moment analysis to its leaves**. 2017. 77p. Monografia (Trabalho de Conclusão de Curso) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2017.

Plant identification is a very old problem for agronomists, agriculturists, biologists and researchers, besides amateurs and the general public. It's a manually done, laborious process, requiring a great amount of time and experience. In this way, it is not an efficient process, wasting a lot of time, especially when the identification of a large set of samples must be made. This work aims develop a computer vision-based method to automatically classify the plants, by analyzing the plant's leaf. The choice of leaf analysis was made because of its greater ease of handling and greater availability, in comparison to other more conventional methods (flower or fruit analysis). The leaves' features are computed by the descriptors of the SIFT (Scale-invariant Feature Transform) algorithm, and by the color moments analysis. Next, the clustering of descriptors is made by the k-means method. The analysis of the frequency of occurrence of these clusters follows a bag-of-words model of representation, which acts as input to a support vector machine (SVM). The machine, already trained with a database, makes the identification of the plant's species. The testing of the developed method allowed a comparison between the performance of the pure SIFT descriptor and the SIFT with color moments descriptor: the first descriptor had its highest success rate at 77,71%, utilizing 500 clusters; the second had its highest success rate at 80,21%, using 300 clusters. The addition of the color moments in the studied cases increased the mean accuracy by 1,75%, with a standard deviation of 4,25%, in relation to the use of only the SIFT descriptor.

Keywords: Computer vision. Plant identification. Leaf analysis. SIFT. Color moments. Bag-of-words. SVM.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

SIFT	*Scale-invariant Feature Transform*
SVM	*Support Vector Machine*
LoG	Laplaciano da Gaussiana
DoG	Diferença das Gaussianas
BoW	*Bag-of-words*
BFF	*Best-Bin-First*
SRM	*Structural Risk Minimization*
USDA	*United States Department of Agricultre*
IDSC	*Inner-distance Shape Context*
SURF	*Speeded Up Robust Features*
SMSD	*Simple and Morphological Shape Descriptors*

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

A identificação de espécies de plantas não é um processo que carrega, em si, algum objetivo maior. Entretanto, tal procedimento é muitas vezes desejável, quando não obrigatório, para o avanço, elucidação e quantificação de experimentos e pesquisas na área das ciências biológicas; além de muitas aplicações mais comuns de gestores de terras e parques, agricultores, diletantes, etc.

A construção de uma base de conhecimento da identidade, distribuição geográfica (característica que varia no tempo) e uso das plantas é essencial tanto para um desenvolvimento sustentável da agricultura quanto para a conservação da biodiversidade (BONNET et al., 2016). Do ponto de vista de gestão de propriedades, a identificação de plantas é importante para saber se determinada planta é ou não intrusiva e se possui algum tipo de risco para o ambiente em questão (MANGOLD, 2013). Além disso, há um constante senso de urgência que impele os cientistas a catalogar a flora da Terra antes que as mudanças climáticas e a expansão urbana deixem alguma marca indelével. É necessário acompanhar e estudar os efeitos dessas mudanças enquanto acontecem para obter um diagnóstico efetivo. (BELHUMEUR et al., 2008).

Em estudos de biodiversidade, entretanto, o "impedimento taxonômico"é sempre presente (GASTON; O'NEILL, 2004). São dificuldades que surgem da organização do corpo de estudo e literatura disponível, adicionando um custo de esforço e tempo considerável para um trabalho de identificação. Destacam-se:

- o viés no conjunto de espécies que foram formalmente descritas (GASTON; O'NEILL, 2004);
- a vasta quantidade de sinônimos existentes, e o grande esforço necessário para eliminá-los (GASTON; O'NEILL, 2004);
- a diminuição gradual de pesquisadores no campo da taxonomia (GASTON; MAY, 1992);
- as dificuldades de tornar-se proficiente em identificação de vários taxa, além do enorme número de espécies que requerem constante verificação (WEEKS et al., 1999);
- a dificuldade no uso de referências adequadas e de terminologias complexas (GASTON; O'NEILL, 2004).

Tais custos, em conjunto com a escassez de profissionais aptos, são repassados para projetos de pesquisa que podem ser postergados, ou até mesmo cancelados, por falta de fundos (GASTON; O'NEILL, 2004). Em suma, a identificação da espécie de plantas é tradicionalmente feita de modo artesanal, consistindo de um processo laborioso e demorado que só pode ser feito por especialistas (OLULEYE et al., 2015).

1.2 Justificativa

A disponibilidade de bases de dados digitais está expandindo rapidamente (SCHMIDT, 2007). As coleções digitais de museus, bibliotecas e universidades abarcam cada vez mais itens. De especial interesse são as digitalizações de herbários, feitas por pesquisadores ou resultados de projetos colaborativos. Tais iniciativas já trouxeram muitos resultados, conforme pode-se ver em (The vPlants Project, 2007) e (C. V. Starr Virtual Herbarium, 2017).

Além da digitalização de herbários já existentes, outras bases de dados sobre plantas, elaborados por cientistas, universidades e institutos de pesquisas, também estão disponíveis digitalmente, resultado da colaboração da comunidade científica internacional. (DAS et al., 2014) é um exemplo de tal base de dados. Nota-se, então, que a disponibilidade e qualidade das bases de dados de plantas é grande, com tendência de crescimento constante.

Esses fatos, aliados com a grande presença no mundo atual de câmeras e computadores, tornam a classificação automática de plantas uma solução não somente viável como desejável (JIN et al., 2015). O problema da identificação de espécies (não somente de plantas) é uma área de crescente interesse em visão computacional (KUMAR et al., 2012), podendo utilizar os recursos mencionados anteriormente de forma extensiva.

(WÄLDCHEN; MÄDER, 2017) realiza uma revisão sistemática dos métodos comumente utilizados em visão computacional para identificação de plantas. A grande maioria de tais métodos baseiam-se em descritores de forma, geralmente aplicados às folhas e flores. As abordagens mais comuns, denominadas como *Simple and Morphological Shape Descriptors* (SMSD), são compostas de medidas de características específicas das folhas das plantas, como sua área ou medidas entre eixos. Embora fossem utilizadas medidas diferentes, variando em complexidade, a grande variação interespécie das folhas de plantas exige métodos mais robustos. Os SMSD simplificam o problema de tal maneira que uma análise significativa torna-se impossível (WÄLDCHEN; MÄDER, 2017). Além disso, em muitos casos a medida desses descritores ainda requer uma decisão do usuário, o que impede uma automação completa do processo.

Em contraste com esses métodos simples, descritores de fronteira ou região são mais adequados para a resolução do problema, podendo ser utilizados em conjunto com os SMSD ou não.

1.3 Objetivos

Nesse contexto, o presente trabalho tem como objetivo desenvolver um método baseado em visão computacional que permita a identificação das plantas através da análise de suas folhas. Em linhas gerais, a folha a qual se quer realizar a identificação vai ser analisada utilizando o algoritmo SIFT e os momentos de cor. Os descritores gerados nesse processo serão agrupados em grupos de características similares através do método *k-means*. Com esses grupos em mãos, pode-se criar uma representação do modelo *bag-of-words* de cada espécie de plantas. Assim, cada espécie será descrita pela frequência de ocorrência de cada palavra visual. Essa representação é, então, usada como entrada pela SVM que, após treinamento com imagens da base de treinamento, realiza a identificação da espécie da folha.

A escolha pela análise das folhas foi feita levando em consideração que, em comparação com outros métodos (análise de flores, frutos ou troncos), as folhas são mais abundantes, tem maior disponibilidade durante o ano e são de manuseio mais fácil. Contudo, também são mais similares entre diferentes espécies (SAITOH; IWATA; WAKISAKA, 2015).

2 CONCEITOS E BASE TEÓRICA

2.1 Pré-processamento de imagens digitais

O pré-processamento da imagem é uma etapa que, com uma imagem entrada pelo usuário, tem como saída, novamente, uma outra imagem, modificada. O objetivo desta etapa é um melhoramento da imagem original, suprimindo ruídos e distorções indesejadas que possam atrapalhar as etapas de processamento seguintes (SONKA; HLAVAC; BOYLE, 1993).

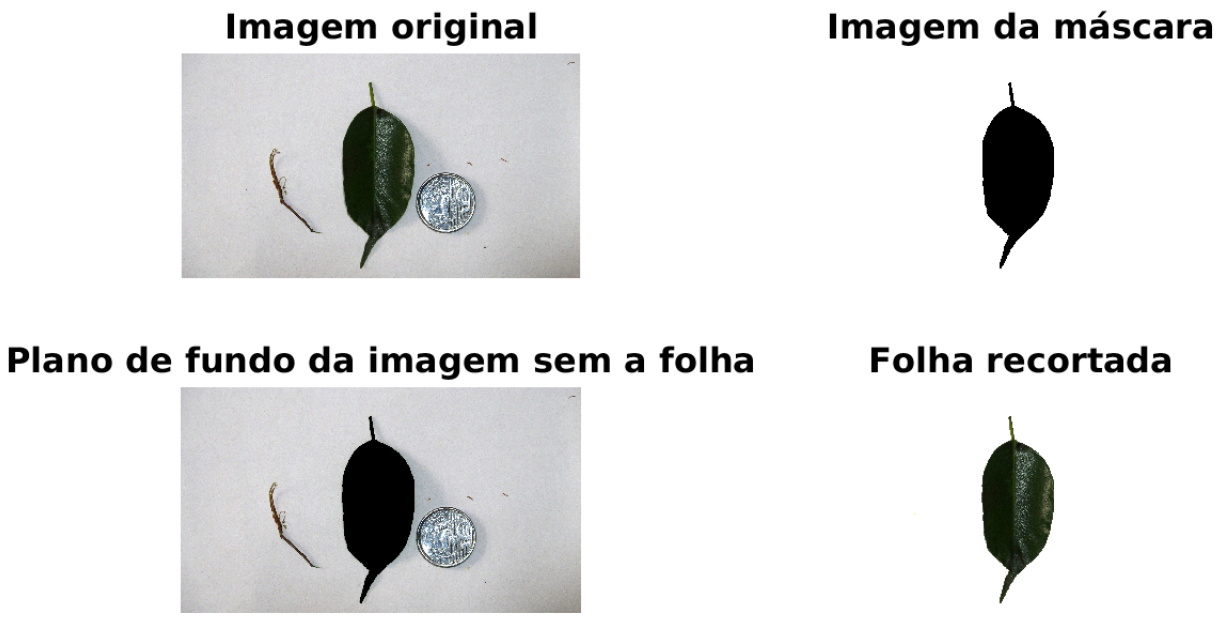
O objetivo do programa é analisar e classificar a folha da planta. Assim, todos os objetos que não sejam parte da folha devem ser removidos, seja fisicamente, antes da captura da imagem, ou posteriormente, por meio de métodos computacionais. Além disso, para efeito de comparação com a base de dados, o plano de fundo da imagem deve ser o mais claro possível, já que a base utilizada disponibiliza as folhas também com um fundo branco (WU et al., 2007). Uma foto da folha da planta sobre uma folha de papel branca ou outra superfície clara é suficiente para a entrada no pré-processamento.

No modelo RGB, cada cor é composta por três cores primárias: vermelho, verde e azul. O padrão atual de armazenamento é de 8 bits por canal de cor (PADMAVATHI; THANGADURAI, 2016). Assim, cada canal passa a ter valores entre 0 e 255, com a intensidade variando de acordo com seu valor. Levando em consideração a distribuição de cor das folhas e o fundo claro em que estão sobrepostas, é possível realizar uma simples filtragem de cores para separar as duas entidades. A cor de fundo possui muito mais componentes no canal azul do que as folhas comumente encontradas (que variam em tons de verde, amarelo e vermelho, com poucos componentes no canal azul). A Figura 1 demonstra o processo em uma imagem com uma folha e outros objetos. Após realizar a extração do plano de fundo, apenas o maior corpo restante (a folha) é selecionado pela máscara.

Nota-se, ainda, que a imagem a ser analisada deverá possuir o mesmo tamanho das imagens disponíveis na base de dados, para permitir uma melhor comparação. Assim, a folha recortada da imagem de entrada é inserida em um fundo branco de tamanho adequado. Caso o recorte for maior que as imagens da base de dados, ele é redimensionado antes dessa etapa.

O algoritmo SIFT utilizado tem como entrada uma imagem em tons de cinza (*grayscale image*). Esse tipo de imagem não contém informações sobre as cores: é apenas uma imagem de intensidades que contém informações sobre a claridade. Em uma imagem comum com armazenamento de 8 bits por pixel, é possível representar 256 tons de cinza, variando do preto até o branco (PADMAVATHI; THANGADURAI, 2016).

Figura 1: Pré-processamento da imagem de uma folha.



Fonte: autoria própria.

A conversão da imagem RGB de entrada para uma imagem em tons de cinza é feita através da seguinte soma com pesos:

$$grayscale = 0.2989R + 0.5870G + 0.1140B \tag{2.1}$$

em que R, G e B são as intensidades do pixel nos respectivos canais vermelhos, verdes ou azuis.

2.2 *Scale-invariant Feature Transform* (SIFT)

Qualquer objeto apresenta várias características, vários pontos de interesse que podem ser extraídos para uma boa análise ou descrição por algoritmos. Além disso, há vários algoritmos que selecionam diferentes grupos de características de maneiras diferentes. O algoritmo SIFT extrai um conjunto de características que não são afetados por muitos dos problemas que causam complicações em outros métodos, como a rotação, mudança de escala ou mudança de iluminação (JAMIL et al., 2015). Isto é, pode-se realizar a análise de um mesmo objeto em múltiplas imagens com diferentes pontos de vista ou iluminações diferentes. O algoritmo SIFT também é bastante resistente ao ruído carregado pela imagem.

Por esses fatores, o algoritmo SIFT foi escolhido para extrair as características das

folhas neste projeto. A descrição aqui feita do algoritmo tem como base os dois artigos de David G. Lowe, criador do algoritmo, (LOWE, 2004) e (LOWE, 1999) e (GONZALES, 2011).

2.2.1 Detecção de extremos nos espaços-escala

A análise dos objetos sempre leva em consideração a escala. Vários algoritmos para a detecção de cantos existem, mas, caso ocorra uma mudança de escala na imagem, alguns cantos podem deixar de existir, como mostra a Figura 2. Neste caso, utilizando a mesma janela computacional, o algoritmo que detecta um canto na primeira imagem não o detecta na segunda, quando ocorre uma mudança de escala. A primeira etapa do algoritmo consiste em achar regiões e características da imagem que são identificáveis independentemente de uma mudança de ponto de vista ou de escala. Isso é atingido utilizando uma função chamada espaço de escala.

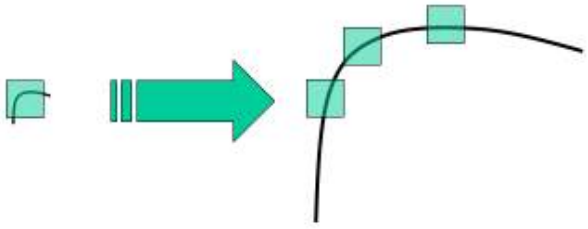
Essa função nada mais é que uma Gaussiana. Sendo $I(x, y)$ uma imagem, fazemos sua transformação para $L(x, y, \sigma)$. A transformação é descrita pela Equação 2.2:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2.2}$$

em que x e y são as coordenadas de cada pixel e σ é o fator de escala. A função Gaussiana, $G(x, y, \sigma)$ é definida como:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+\frac{y^2}{2\sigma^2})} \tag{2.3}$$

Figura 2: Detecção de canto em diferentes escalas. Utilizando a mesma janela computacional, o canto só é detectado em uma escala.



Fonte: adaptada de (OPENCV, 2017).

Como se pode ver, a transformação é dependente do fator de escala, σ. Na Figura 3, pode-se ver o resultado da transformação para diferentes escalas.

A aplicação sucessiva da convolução por Gaussianas com diferentes fatores de escala leva a criação de um espaço de escalas. Cada objeto desse espaço é a imagem original convolucionada em uma escala diferente. No SIFT, a geração desse espaço de escalas é feita

Figura 3: Resultado da transformação para $\sigma = 1$, 2, 4, 8, 16 e 32. O resultado da convolução pela Gaussiana é um borramento da imagem, dependente da escala.

Fonte: adaptada de (SINHA, 2017).

de um modo iterativo: a imagem original é convolucionada pela Gaussiana em diferentes escalas. Em seguida, ela é dividida pela metade e o mesmo processo acontece. O processo se repete para cada oitava assim gerada. A decisão do número de oitavas utilizadas varia com a aplicação, mas em (LOWE, 2004), é recomendado quatro oitavas e cinco convoluções. A Figura 4 exemplifica o processo.

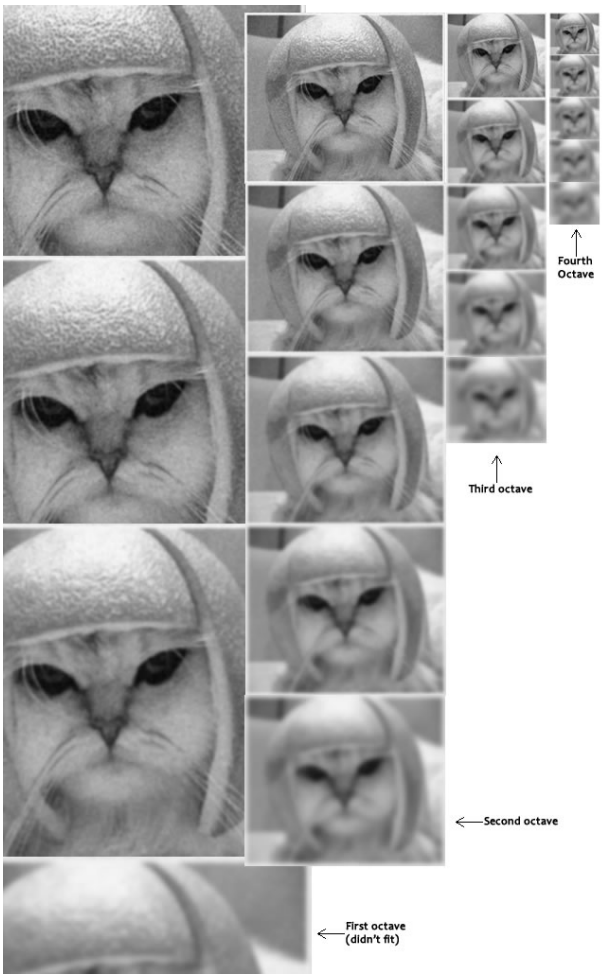
O próximo passo é aplicar o operador Laplaciano (Δ^2) às imagens $I(x, y)convolucionadaspelaGaussi$
y), gerando o Laplaciano da Gaussiana (LoG). Isto é, calcular a Equação 2.4:

$$\Delta^2[G(x, y) * I(x, y)] \tag{2.4}$$

A derivada de segunda ordem é sensível ao ruído. Quando aplicada a uma Gaussiana, a entrada já é uma imagem borrada, o que diminui a influência do ruído. O LoG é uma operação que localiza as bordas e cantos da imagem, características que auxiliam na busca pelos pontos-chave ideais. Em outras palavras, acha-se os máximos locais em um determinado espaço-escala. Esses máximos são indicadores da presença ou não de um ponto-chave em uma determinada escala.

O cálculo do LoG, entretanto, tem um alto custo computacional. A solução é uma

Figura 4: Geração do espaço de escalas.



Fonte: adaptada de (SINHA, 2017).

aproximação do LoG através da Diferença das Gaussianas (DoG). A função DoG é definida por:

$$DoG = G(x, y, k\sigma) - G(x, y, \sigma) \tag{2.5}$$

em que x e y são pontos da imagem, k é uma constante que varia a escala, σ é o fator de escala e $G(x, y, \sigma)$ é a Gaussiana.

Simplificando:

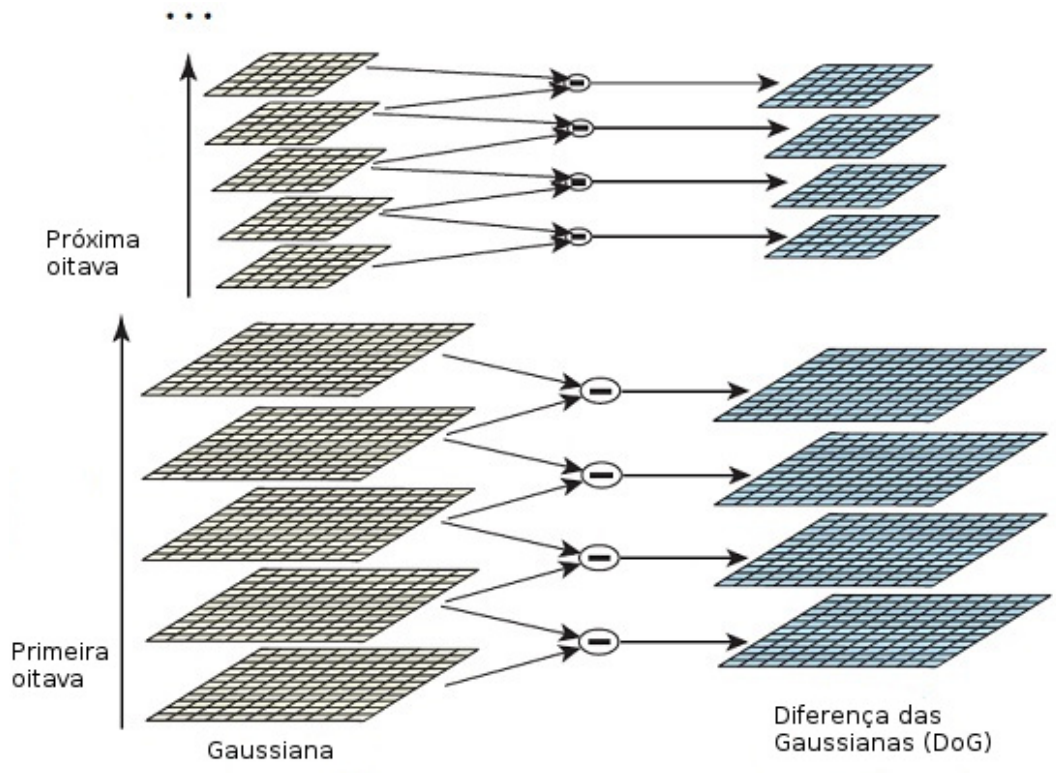
$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \tag{2.6}$$

$$D(x, y, \sigma) = G(x, y, k\sigma) * I(x, y) - G(x, y, \sigma) * I(x, y) \tag{2.7}$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{2.8}$$

em que $D(x, y, \sigma)$ é a imagem transformada pela DoG e $L(x, y, \sigma)$ é a imagem original transposta para uma escala σ no espaço de escala. Assim, o cálculo da DoG é feito pela simples subtração de imagens borradas pela Gaussiana em diferentes, mas próximas, escalas. A Figura 5 ilustra o processo. O resultado são imagens onde detalhes indesejados e ruídos são eliminados, enquanto as características fortes são realçadas.

Figura 5: Cálculo da DoG.



Fonte: adaptada de (SINHA, 2017).

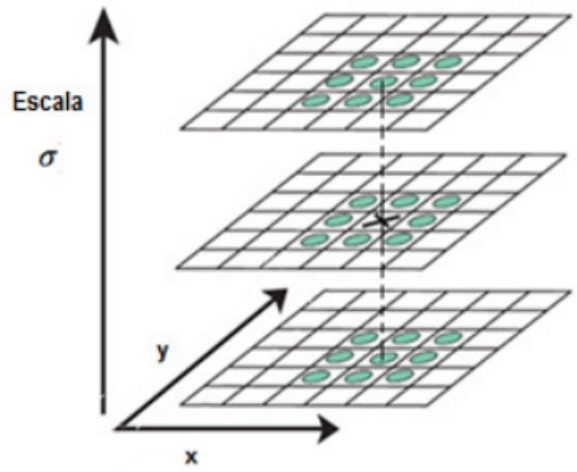
Com a DoG em mãos, é feito a busca por máximos nos espaços-escala. A comparação é feita como ilustra a Figura 6. Um pixel na imagem é comparado com seus oito vizinhos no mesmo espaço e com 9 pixels nas escalas superiores e inferiores. Se esse ponto for um máximo local, é um candidato a ponto-chave.

2.2.2 Localização de pontos-chave

Todos os pontos detectados como máximos no passo anterior são candidatos a pontos-chave. Entretanto, sua localização está sendo feita apenas a nível de pixels. A coincidência de um ponto de extremo em um determinado pixel é rara. Assim, é necessário aumentar a precisão da localização dos extremos para subpixels. Isso é feito por meio da da Equação 2.9:

$$D(\bar{x}) = D + \frac{\partial D^T}{\partial \bar{x}}\bar{x} + \frac{1}{2}\bar{x}^T\frac{\partial^2 D^T}{\partial \bar{x}^2}\bar{x} + ... \tag{2.9}$$

Figura 6: Comparação entre os pixels para a verificação de máximo local.



Fonte: adaptada de (GONZALES, 2011).

em que $D(x)$ é a DoG, calculada no ponto de amostragem $\bar{x} = (x, y, \sigma)^T$. A Equação 2.9 nada mais é que a expansão por séries de Taylor da DoG aplicada a uma imagem, deslocada tal que sua origem seja o ponto de amostragem.

(LOWE, 2004) recomenda a adição de um limiar para a intensidade dos máximos calculados por este método. Se a intensidade for menor do que 0.03, o ponto é rejeitado. Além disso, a função DoG é bastante sensível à bordas. Então, as bordas precisam ser eliminadas. Isto é feito primeiramente utilizando uma matriz hessiana 2x2 para computar a curvatura principal:

$$H(x, y) = \begin{bmatrix} D_{xx} & D_{xy} \\ Dxy & D_{yy} \end{bmatrix} \tag{2.10}$$

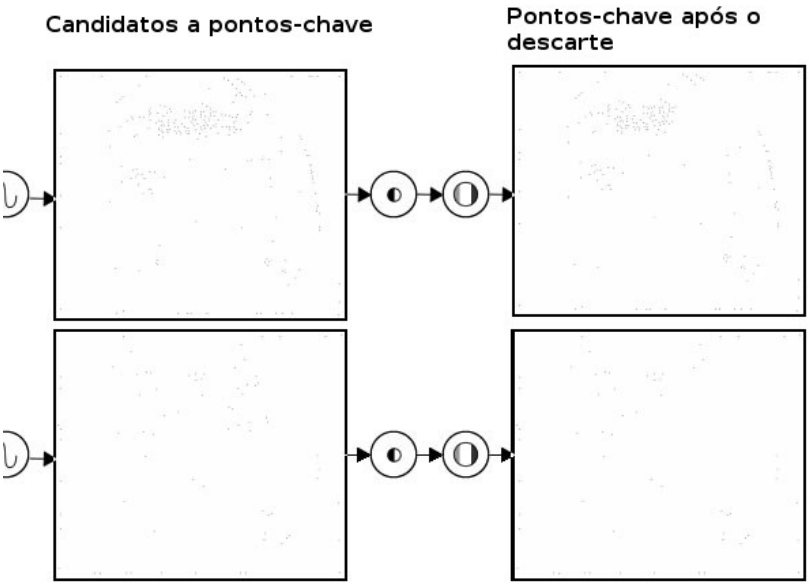
em que $H(x, y)$ é a matriz Hessiana, D_{xx}, D_{xy} e D_{yy} são as derivadas parciais de segunda ordem da DoG.

Para bordas, um autovalor é maior do que o outro. Então, a filtragem é feita pela equação 2.11:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r} \tag{2.11}$$

em que r é a razão entre os autovalores da matriz H. Assim, pode-se ajustar o limiar ao escolher qual r é aceito. A Figura 7 demonstra, à esquerda, os candidatos a pontos-chave de duas escalas diferentes do espaço-escala da imagem da Figura 4. Após realizar os procedimentos mencionados anteriormente, alguns pontos são descartados, exibidos à esquerda.

Figura 7: Exemplo de filtragem de bordas pelo processo descrito.



Fonte: adaptada de (SINHA, 2017).

2.2.3 Atribuição de Orientação

A cada ponto-chave escolhido nos processos anteriores é atribuído uma orientação. Como consequência, o descritor torna-se invariante à rotação.

Para cada imagem $L(x, y)$ em um espaço-escala, são calculados os seguintes parâmetros em torno dos pontos-chave:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \tag{2.12}$$

$$\theta(x,y) = tan^{-1}\frac{(L(x,y+1) - L(x,y-1))}{(L(x+1,y) - L(x+1,y))} \tag{2.13}$$

em que $m(x, y)$ é a magnitude e $\theta(x, y)$ é a orientação de um determinado ponto (x, y) de uma imagem do espaço-escala $L(x, y)$.

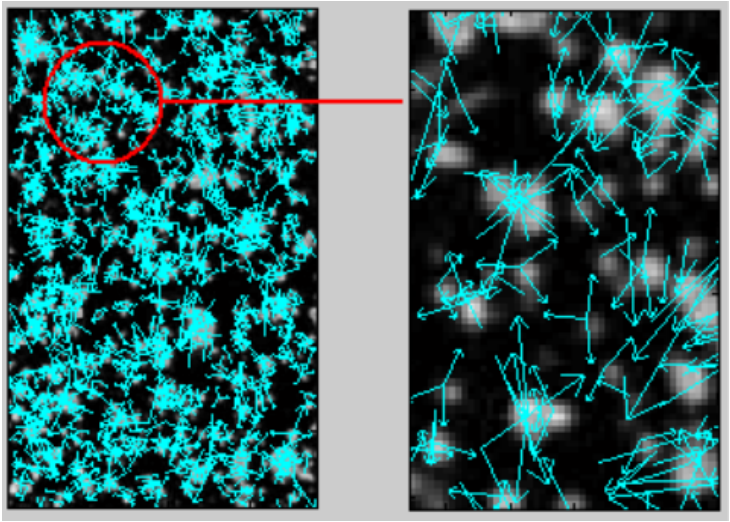
Após isso, um histograma das orientações é criado. No histograma, os 360 graus de orientação são particionados em 36 categorias. Cada ponto na vizinhança de um ponto-chave é adicionado ao histograma de acordo com parâmetros com peso. O primeiro peso é a magnitude. O segundo peso é uma janela circular Gaussiana com σ igual a 1.5 vezes o tamanho da escala do ponto-chave. A janela pode ser descrita pela Equação 2.14:

$$g(\Delta x, \Delta y, \sigma) = \frac{1}{2\pi\sigma^2} e^{(\Delta x^2 + \Delta y^2)/2\sigma^2} \tag{2.14}$$

em que $g(\Delta x, \Delta y, \sigma)$ é a janela, Δx e Δy são as distâncias entre os pontos verificados e o ponto-chave, x e y são pontos da imagem e σ é o fator de escala.

Os picos do histograma correspondem a direções dominantes nos gradientes locais. São considerados como picos qualquer intensidade que chegue até 80% do valor máximo. No caso de múltiplos picos, o ponto-chave receberá múltiplas orientações, facilitando sua futura identificação. Pode-se ainda interpolar os valores do histograma mais próximos ao pico, para ampliar a exatidão do procedimento. A Figura 8 ilustra o resultado da adição da orientação para os pontos-chave do modo descrito nessa seção.

Figura 8: Exemplo de adição dos vetores de magnitude e orientação aos pontos chaves.



Fonte: adaptada de (GONZALES, 2011).

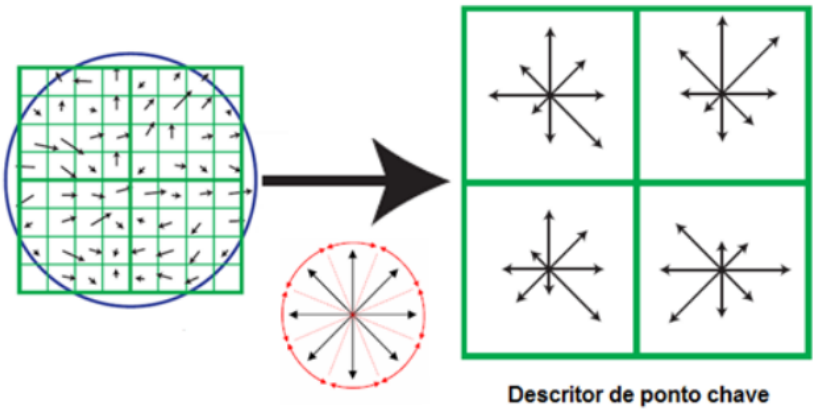
2.2.4 Geração do descritor aos pontos-chave

Nos passos anteriores, a invariância a escalas e rotação foi adicionada. Agora, cada ponto-chave receberá um descritor invariante a iluminação e ponto de vista 3D. Os passos a seguir são realizados com os valores de magnitude e orientação da seção anterior normalizados.

O descritor é criado pelo cálculo da magnitude e orientação dos gradientes em torno do ponto-chave. Uma função Gaussiana é utilizada para parametrizar a magnitude do gradiente dos pontos com um peso. O parâmetro σ é igual a metade da largura da janela do descritor. Essa parametrização evita reações altas no descritor conforme ocorre a varredura da janela. Isto é, a função Gaussiana suaviza a janela. O processo descrito está ilustrado na Figura 9.

Entretanto, as variações de luminosidade podem acabar gerando dois descritores extremamente diferentes para um mesmo objeto. Para atingir a invariância à iluminação, os descritores são normalizados.

Figura 9: Geração do descritor.



Fonte: adaptada de (GONZALES, 2011).

Quando ocorrem mudanças homogêneas de brilho, levando em conta que tais mudanças afetam todos os pixels da imagem do mesmo modo, a normalização mantém a invariância. O mesmo pode-se dizer quanto às mudanças homogêneas de contraste.

Já as variações não lineares, causadas por sensores de câmeras ou reflexão de superfícies, possuem alta influência na magnitude, mas não na orientação dos descritores. O efeito pode ser reduzido aplicando-se um limiar após a normalização. (LOWE, 2004) recomenda um limiar de 0.2. Isto quer dizer que a influência das magnitudes não é tão importante quanto a influência das orientações.

2.3 Momentos de cor

A análise probabilística da distribuição de cores em uma imagem pode ser realizada na medição de seus momentos de cor. Essa análise é tipicamente utilizada em aplicações em que se quer comparar imagens: uma nova imagem é comparada com uma base de dados (YU et al., 2002). Os momentos de cor retornam informações sobre a similaridade das imagens. Pode-se computar momentos de cor para qualquer representação de cor escolhida. Para uma imagem RGB, os momentos de cor de cada canal são calculados separadamente. Eles são computados da mesma maneira que os momentos em uma distribuição probabilística.

Além disso, os momentos de cor são invariantes à rotação e mudança de escala, ampliando o seu alcance de aplicações (YU et al., 2002). Outra vantagem é o fato de, ao se armazenar os momentos de cor, não ser necessário armazenar a distribuição de cor completa: apenas os momentos bastam (STRICKER; ORENGO, 1995).

2.3.1 Primeiro momento de cor: média

O primeiro momento de cor é a média. Representa a média de cor na imagem. Pode ser calculado como:

$$\bar{Y} = \sum_{i=1}^{N} \frac{1}{N} Y_i \tag{2.15}$$

em que $\bar{Y}$ é o valor da média, N é o número de pixels da imagem e Y_i é o valor do pixel da imagem I no ponto (x, y), isto é, $I(x, y)$.

2.3.2 Segundo momento de cor: desvio padrão

O desvio padrão é simplesmente a raiz quadrada da variância da distribuição de cor. Pode ser calculado como:

$$\sigma = \sqrt{\left(\frac{1}{N} \sum_{i=1}^{N} (Y_i - \bar{Y})^2\right)} \tag{2.16}$$

em que σ é o desvio padrão, N é o número de pixels da imagem e Y_i é como antes.

2.3.3 Terceiro momento de cor: obliquidade

O terceiro momento de cor é a obliquidade. Ele dá informações sobre a simetria da distribuição. Pode ser calculado como:

$$o = \frac{\sum_{i=1}^{N} \frac{(Y_i - \bar{Y})^3}{N}}{\sigma^3} \tag{2.17}$$

em que em que o é a obliquidade, N é o número de pixels da imagem, σ é o desvio padrão e Y_i é como antes.

2.3.4 Quarto momento de cor: curtose

O quarto momento de cor é a curtose. Ela compara o formato da distribuição analisada com o formato da distribuição normal. Pode ser calculada como:

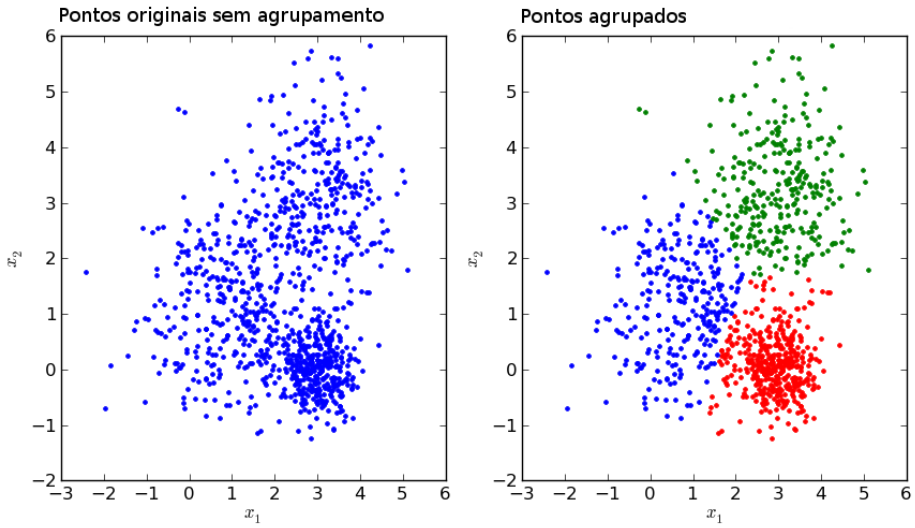
$$c = \frac{\sum_{i=1}^{N} \frac{(Y_i - \bar{Y})^4}{N}}{\sigma^4} \tag{2.18}$$

em que em que c é a curtose, N é o número de pixels da imagem, σ é o desvio padrão e Y_i é como antes.

2.4 Método *k-means*

O método *k-means* é um método de agrupamento de dados não-supervisionado que particiona um grupo de observações em grupos com características similares. A distribuição de cada dado para um grupo em específico é feita com base na comparação da média: o dado vai para o grupo com média mais similar. É um método já estabelecido e com várias aplicações em identificações de espécies (SABRI; IBRAHIM; ROSMAN, 2016), já que pode agrupar grandes bases de dados rapidamente (VALLIAMMAL; GEETHALAKSHMI, 2012). A Figura 10 exemplifica o agrupamento de um conjunto de dados em três grupos distintos.

Figura 10: Agrupamento de dados pelo método *k-means*.



Fonte: adaptada de (PYPR, 2010).

Existem várias implementações do algoritmo *k-means*. A mais comum baseia-se no algoritmo de Lloyd. O algoritmo tem por fim minimizar a função de erro

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2 \tag{2.19}$$

em que $\left\| x_i^j - c_j \right\|^2$ é a distância entre o ponto analisado e o centroide do grupo. x_i é um ponto do conjunto de dados que se quer agrupar e c_j a centroide correspondente. Isso ocorre através dos seguintes passsos:

1. escolha dos centroides iniciais;
2. cálculo das distâncias de todos os pontos analisados até os centroides;
3. agrupamento dos dados nos centroides mais próximas;

4. cálculo da média de cada grupo para obtenção dos novas centroides;

5. repetir passos até que os centroides não se movam mais.

O algoritmo de Elkan é uma variação do algoritmo de Lloyd (ELKAN, 2003). Ele baseia-se na ideia de que, como os centroides não se deslocam muito após algumas iterações, então a maioria dos cálculos da distância do ponto ao centro podem ser evitadas. A detecção de quais distâncias não precisam ser computadas é feita através de uma estimativa que utiliza a desigualdade triangular para limitar inferiormente e superiormente as distâncias após a atualização dos centroides.

Matematicamente, tem-se

$$\|x_i - c_{q_i}\|_p \leq \|c - c_{q_i}\|_p/2 \quad \Rightarrow \quad \|x_i - c_{q_i}\|_p \leq \|x_i - c\|_p. \tag{2.20}$$

Isto é, se a distância entre o ponto x_i e seu centroide c_{q_i} for menor do que metade da distância do centroide c_{q_i} até outro centroide c, então c pode ser omitida quando da busca por um novo centroide para x_i.

Caso um centroide c for atualizado para $\hat{c}$, então a distância de x até $\hat{c}$ está limitada de acordo com a equação 2.21

$$\|x - c\|_p - \|c - \hat{c}\|_p \leq \|x - \hat{c}\|_p \leq \|x - \hat{c}\|_p + \|c + \hat{c}\|_p. \tag{2.21}$$

O método de escolha dos centroides inciais também pode ser otimizado através do *k-means++* (ARTHUR; VASSILVITSKII, 2007). Nesse caso, uma nova etapa de inicialização de centroides é realizada, para somente depois prosseguir com o algoritmo *k-means* normal.

Essa etapa consiste em:

- Escolher um centroide aleatoriamente entre os pontos x_1, ..., x_n;
- para cada ponto x, a distância entre ele e a mais próxima centroide já escolhida é calculada;
- uma nova centroide é escolhida, novamente aleatoriamente, mas cada ponto possui probabilidade proporcional ao quadrado da distância calculada;
- os passos anteriores são repetidos até que o número de centroides desejadas seja atingido

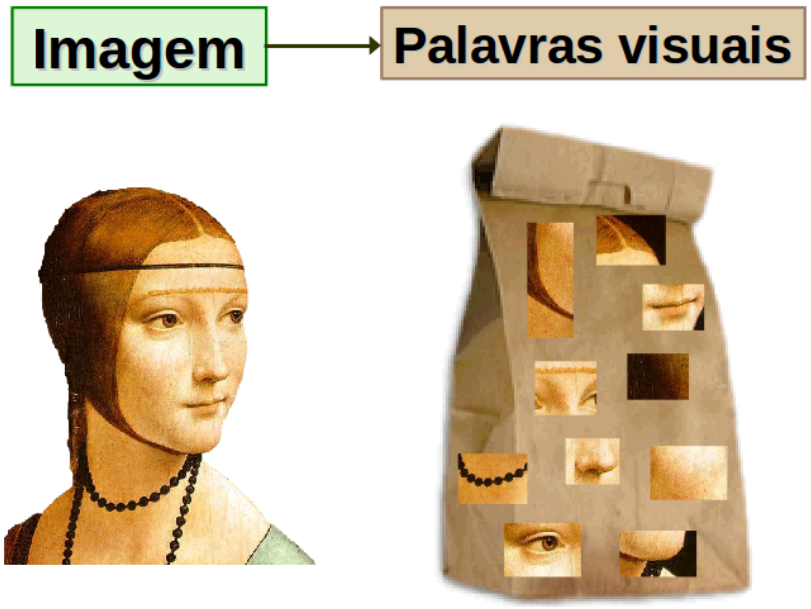
Após esses passos, o algoritmo *k-means* normal pode ser continuado.

Apesar da seleção inicial de centroides atrasar um pouco as etapas iniciais do algoritmo, ele faz com que a etapa normal do *k-means* convirja rapidamente, apresentando um tempo de computação mais baixo (ARTHUR; VASSILVITSKII, 2007).

2.5 *Bag-of-words*

O modelo *bag-of-words* é um método utilizado tradicionalmente no processamento de linguagens naturais. Neste caso, o objeto que se quer analisar é algum texto. Ele é analisado e uma representação baseada na frequência de aparição de cada palavra é criada: cada texto é composto por diferentes palavras com diferentes frequências. Tal representação não é ordenada. A partir da comparação entre as frequências de palavras que os textos possuem, seus histogramas, é possível traçar conclusões sobre sua similaridade (SIVIC; ZISSERMAN, 2009).

Figura 11: Representação por palavras visuais.

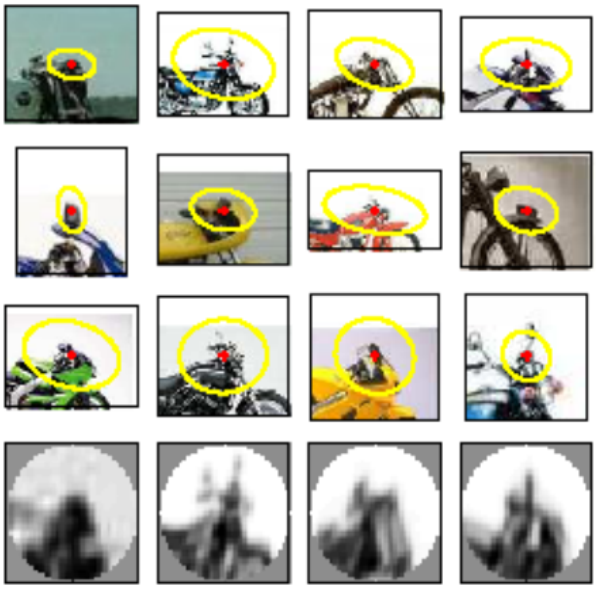


Fonte: adaptado de (FEI-FEI; FERGUS; TORRALBA, 2009).

O mesmo modelo pode ser adaptado para visão computacional (SIVIC; ZISSERMAN, 2009). Neste caso, as imagens são analisadas em busca de componentes visuais significantes, assim como as palavras em um texto. São chamadas de *palavras visuais* (ZHANG; JIN; ZHOU, 2010). A Figura **??**, adaptada de (FEI-FEI; FERGUS; TORRALBA, 2009) demonstra uma partição com alguns componentes. O método para a criação de tal modelo são descritos abaixo.

A análise da imagem é feita através de um descritor – por exemplo, o algoritmo SIFT. Após esse passo, cada característica detectada na imagem pelo descritor é representada

Figura 12: Agrupamento de imagens similares em um dicionário.



Fonte: adaptada de (FEI-FEI; FERGUS; TORRALBA, 2009).

por um vetor de mesmo tamanho. A imagem, então, é descrita pelo conjunto desses vetores. Aqui, a ordem desses vetores não é importante e, nesse momento, não há qualquer agrupamento deles.

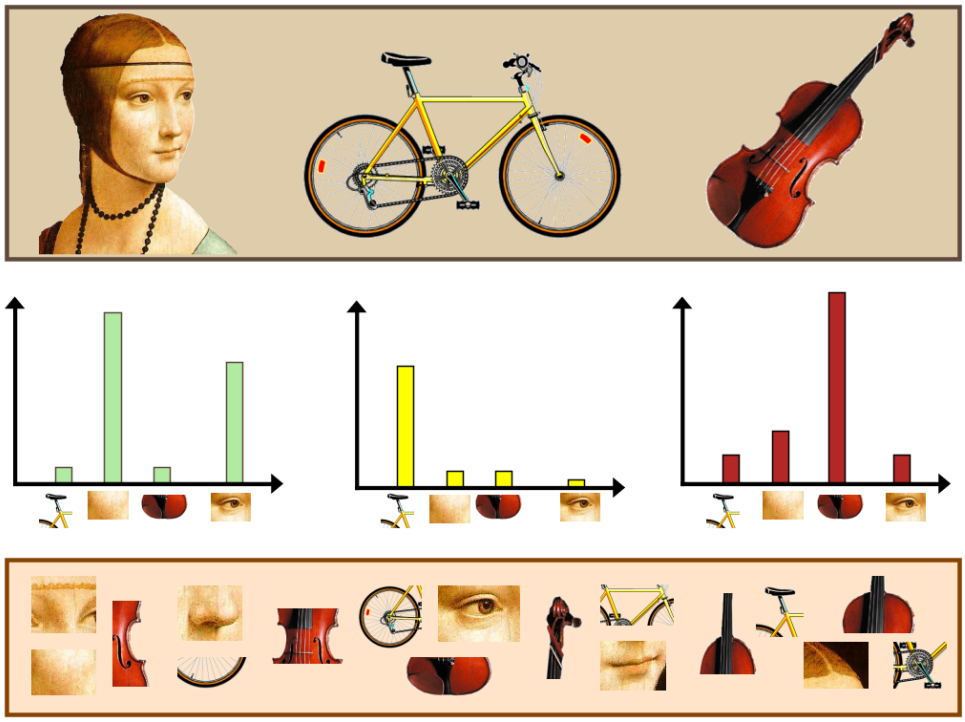
O próximo passo é o agrupamento dos vetores de características similares para a criação de um dicionário de palavras visuais. Isso pode ser feito por vários métodos: *k-means*, kd-tree, KNN. Com as palavras visuais agrupadas, pode-se realizar a análise de frequência de palavras de uma imagem. Isso é exemplificado na Fig 12. Ou seja, cada imagem pode ser descrita por meio da contribuição relativa que cada palavra visual possui na imagem.

Assim, o resultado final do processamento da imagem é um histograma representando a frequência que cada palavra visual, descrita e agrupada anteriormente, possui na imagem original. A partir da comparação de um histograma de uma imagem em específico com um conjunto de histogramas de uma base de dados, é possível a identificação e classificação baseada em suas similaridades. A Figura 13 representa o processo da análise da distribuição de frequências de palavras visuais pelo histograma.

2.6 *Support Vector Machine* (SVM)

Uma Máquina de Vetores de Suporte (do inglês *Support Vector Machine*) é um algoritmo de aprendizado de máquinas supervisionado bastante utilizado na resolução de problemas que envolvem tanto classificação de dados quanto regressão (LORENA; CARVALHO, 2007). O desempenho das SVMs é igual ou superior a outros métodos de

Figura 13: Processo de análise da distribuição das palavras visuais pela análise do histograma.



Fonte: adaptada de (FEI-FEI; FERGUS; TORRALBA, 2009).

classificação por redes neurais artificias (LORENA; CARVALHO, 2007).

Em linhas gerais, as SVMs são construídas seguindo o princípio da Minimização do Risco Estrutural (SRM, do inglês *Structural Risk Minimization*). Este princípio evita o sobreajuste ao equilibrar a complexidade de seu modelo contra sua precisão na classificação dos dados.

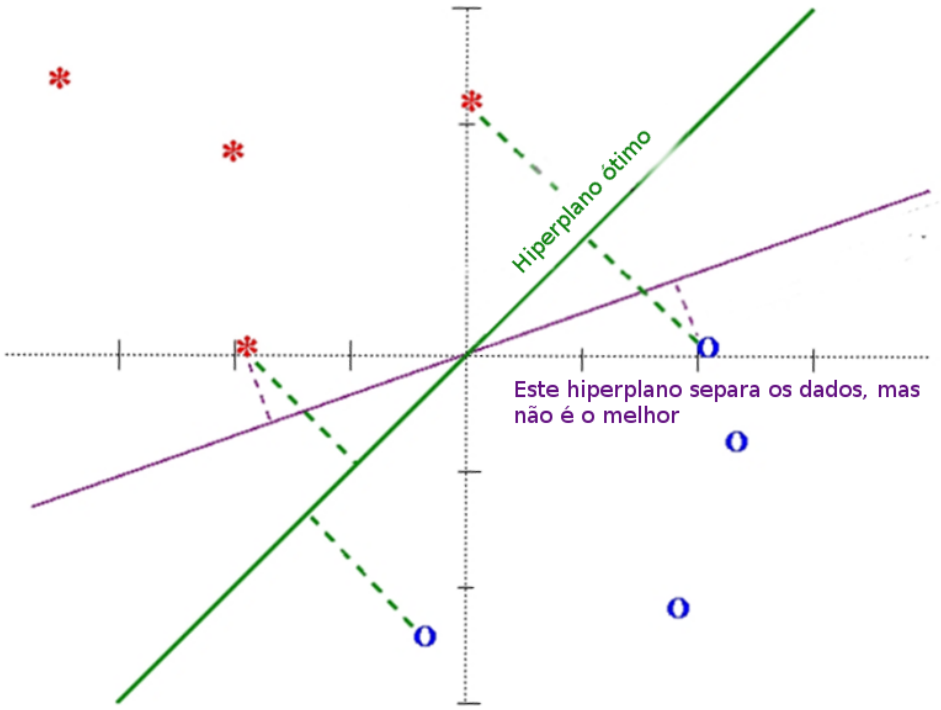
A SVM recebe um conjunto de dados de treinamento e retorna um hiperplano que melhor os categoriza. Este hiperplano é o limite de decisão: divide o conjunto de dados em grupos. O hiperplano escolhido pela SVM é aquele que maximiza a sua distância ao elemento mais próximo de cada categoria, como pode-se ver na Figura 14.

Caso a entrada da SVM sejam dados linearmente separáveis, o hiperplano será, também, uma reta, não criando maior complexidade. Em caso de dados não lineares, ainda é possível a classificação pela SVM, embora de uma forma mais complexa.

Nesses casos, a análise do limite de decisão é feita em uma nova dimensão, criada de forma que o cálculo do hiperplano limite seja mais conveniente, de maneira análoga à classificação de dados linearmente separáveis. O limite de decisão obtido nessa nova dimensão é, então, mapeado de volta às dimensões anteriores.

Contudo, as transformações exigidas por esse método tendem a deixar o algoritmo

Figura 14: Comparação entre um hiperplano que separa as categorias e o hiperplano ótimo.



Fonte: adaptada de (BOSWELL, 2002).

exigente quanto a recursos computacionais. Quanto mais dimensões e transformações envolvidas no cálculo, maior o custo computacional. A aplicação dos mapeamentos para cada vetor na base de dados tornaria o algoritmo inviável.

Na prática, a classificação por uma SVM não necessita da análise integral de cada vetor da base. É necessário apenas conhecer o produto escalar. Assim, ao invés de mapear os vetores para uma nova dimensão, tudo que é preciso fazer é conhecer as regras do produto escalar nessa dimensão. A não necessidade dos mapeamentos dos vetores para outras dimensões, em função da adoção do cálculo do produto escalar é chamado de *truque do kernel*.

A SVM como foi apresentada é geralmente utilizada para a classificação de casos binários, em que há apenas duas classes. A expansão para problemas com classes múltiplas, entretanto, não é de alta complexidade. Basta reduzir o problema de classes múltiplas em uma sequência de problemas de classes binárias.

3 MATERIAIS E MÉTODOS

3.1 Base de dados

A base de dados de folhas de plantas utilizada foi a disponibilizada pelo projeto FLAVIA (WU et al., 2007). Essa base de dados é composta de 1907 imagens das folhas de 32 espécies diferentes de plantas; cada espécie possui em torno de 50 imagens. Cada imagem possui dimensões de 1600x1200. As imagens foram agrupadas de diferentes fontes, como repositórios universitários, repositórios da USDA ou da Wikipédia. Uma lista completa encontra-se em (WU et al., 2007). A Figura 15 exibe algumas folhas da base de dados.

Como a quantidade de imagens disponíveis varia entre as diferentes espécies, a menor quantidade de fotos disponíveis por espécie foi escolhida como padrão. A espécie de número 11 possui o menor número de imagens: 50. Assim, dentre as imagens disponíveis para as outras espécies, foram selecionadas 50, fazendo com que o número de imagens por espécies se tornasse compatível.

Realizada essa normalização, dentre as 1907 imagens disponíveis, 1600 imagens foram utilizadas no programa. Dessas, 1120 (70%) foram utilizadas para treinamento, totalizando 35 imagens por espécies. Os as 480 imagens restantes (30%) foram utilizados para validação e testes do programa. A tabela 1 mostra a relação entre o número de imagens utilizadas para treinamento e teste, para cada espécie de planta disponível na base de dados. A tabela 1 também ordena as espécies e atribui um número de identificação para cada espécie.

3.2 Método proposto

Em linhas gerais, o método proposto para a identificação de plantas recebe como entrada uma imagem do banco de dados, conforme exposto anteriormente. Em seguida, a análise da imagem é feita por meio tanto do algoritmo SIFT quanto dos momentos de cor. O próximo passo é o agrupamentos dos descritores gerados em grupos de características similares, o que é feito através do método *k-means*, formando as palavras visuais. Os grupos gerados nesse passo formam a base do modelo *bag-of-words* de representação, onde cada imagem é descrita a partir de um histograma da frequência relativa de cada palavra visual. Por fim, o histograma gerado serve como entrada para a SVM. O fluxograma do método desenvolvido encontra-se na Figura 16.

A Figura 16 mostra uma divisão no fluxograma. Embora utilizem passos similares, há uma divisão entre as etapas de treinamento e validação do método.

À direita, estão os passos utilizados no treinamento da SVM. Nesses passos são

Tabela 1: Porcentagem e número das imagens utilizadas para treinamento e testes.

Nº	Nome da espécie	Nº de imagens	Para treinamento	Para teste
1	Phyllostachys edulis (Carr.) Houz.	59	35 (59,32%)	15 (25%)
2	Aesculus chinensis	63	35 (55,55%)	15 (23%)
3	Berberis anhweiensis Ahrendt	72	35 (48,61%)	15 (20%)
4	Cercis chinensis	73	35 (47,94%)	15 (20%)
5	Indigofera tinctoria L.	56	35 (62,5%)	15 (26%)
6	Acer Palmatum	62	35 (56,45%)	15 (24%)
7	Phoebe nanmu (Oliv.) Gamble	52	35 (67,3%)	15 (28,%)
8	Kalopanax septemlobus (Thunb. ex A.Murr.) Koidz.	59	35 (59,32%)	15 (25%)
9	Cinnamomum japonicum Sieb.	55	35 (63,63%)	15 (27%)
10	Koelreuteria paniculata Laxm.	65	35 (53,84%)	15 (23%)
11	Ilex macrocarpa Oliv.	50	35 (70%)	15 (30%)
12	Pittosporum tobira (Thunb.) Ait. f.	63	35 (55,55%)	15 (23%)
13	Chimonanthus praecox L.	52	35 (67,3%)	15 (28,%)
14	Cinnamomum camphora (L.) J. Presl	65	35 (53,84%)	15 (23%)
15	Viburnum awabuki K.Koch	60	35 (58,33%)	15 (25%)
16	Osmanthus fragrans Lour.	56	35 (62,5%)	15 (26%)
17	Cedrus deodara (Roxb.) G. Don	77	35 (45,45%)	15 (19%)
18	Ginkgo biloba L.	62	35 (56,45%)	15 (24%)
19	Lagerstroemia indica (L.) Pers.	61	35 (57,37%)	15 (24%)
20	Nerium oleander L.	66	35 (53,03%)	15 (22%)
21	Podocarpus macrophyllus (Thunb.) Sweet	60	35 (58,33%)	15 (25%)
22	Prunus serrulata Lindl. var. lannesiana auct.	55	35 (63,63%)	15 (27%)
23	Ligustrum lucidum Ait. f.	55	35 (63,63%)	15 (27%)
24	Tonna sinensis M. Roem.	65	35 (53,84%)	15 (23%)
25	Prunus persica (L.) Batsch	54	35 (64,81%)	15 (27%)
26	Manglietia fordiana Oliv.	52	35 (67,3%)	15 (28,%)
27	Acer buergerianum Miq.	53	35 (66,03%)	15 (28,%)
28	Mahonia bealei (Fortune) Carr.	55	35 (63,63%)	15 (27%)
29	Magnolia grandiflora L.	57	35 (61,4%)	15 (26%)
30	Populus ×canadensis Moench	64	35 (54,68%)	15 (23%)
31	Liriodendron chinense (Hemsl.) Sarg.	53	35 (66,03%)	15 (28,%)
32	Citrus reticulata Blanco	56	35 (62,5%)	15 (26%)

definidas centroides que vão ser utilizadas para o agrupamento das palavras visuais. A etapa de classficação ocorre à esquerda, onde não são geradas novas centroides, e sim utilizadas as geradas pelo processo de treinamento. Em ambos os casos, a análise da imagem da folha da planta é feita do mesmo modo, através do SIFT e dos momentos de cor.

Além disso, para a aferição do impacto dos momentos de cor no descritor final, a implementação do método foi realizada em duas etapas diferentes: a primeira utilizando apenas o descritor SIFT puro e a segunda utilizando tanto o descritor SIFT quanto os momentos de cor.

Nos pontos a seguir, o funcionamento de cada bloco do fluxograma é aprofundado:

- Entrada da imagem da folha a ser analisada. Pré-processamento da imagem.

A imagem da folha a ser analisada é enviada para o programa. É feito, então, um

Figura 15: Algumas folhas da base de dados do projeto FLAVIA.

Fonte: adaptada de (WU et al., 2007).

pré-processamento da imagem antes da extração das características. Nessa parte, a imagem é redimensionada, uma máscara da folha é criada e os objetos indesejados são retirados antes da análise da folha.

- Extração das características das folhas feita pelo algoritmo SIFT (*Scale-invariant Feature Transform)*.

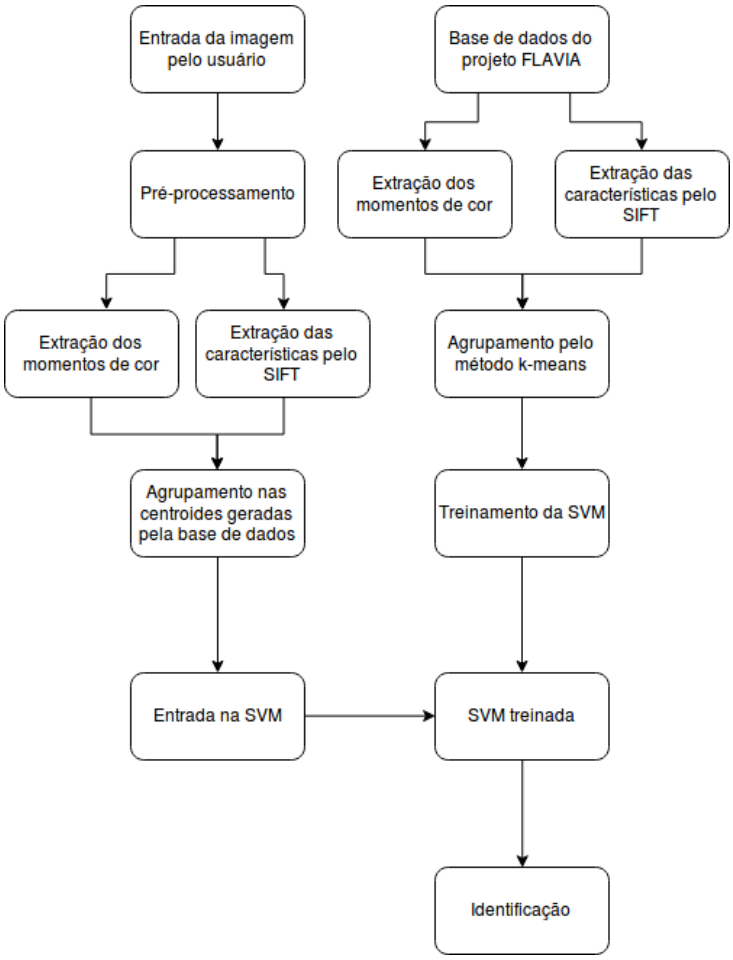
O algoritmo SIFT é invariante não somente à mudança de escala, mas também quanto à rotação, iluminação e ponto de vista, características que o tornam um bom candidato para a extração das características das folhas (LOWE, 2004).

- Extração dos momentos de cor.

A análise por momentos de cor realiza um estudo probabilístico sobre a distribuição de cores na imagem. Como o SIFT, também é invariante à rotação e à mudança de escala (YU et al., 2002).

- Divisão dos descritores em grupos através do método *k-means*. Representação por palavras visuais.

Figura 16: Fluxograma do método proposto.



Fonte: autoria própria.

Os descritores obtidos pelo SIFT na análise da base de dados são quantizados pelo método *k-means*, que é um método simples, eficaz e comumente utilizado (PATIL et al., 2016). Cada grupo gerado nesse método, de acordo com o modelo *bag-of-words*, pode ser considerado como uma palavra visual da imagem.

- Criação do histograma de ocorrência das palavras visuais.

Cada imagem é novamente analisada e comparada com as palavras visuais geradas no passo anterior. A frequência de ocorrência, em cada imagem, de cada palavra visual é gravada em um histograma.

- Treinamento da SVM (*Support Vector Machine*) com histogramas gerados. Comparação da folha enviada com a base de dados. Identificação da espécie da folha.

A identificação é feita através de uma SVM treinada com as informações de uma base de dados já processada. Tal processamento é feito de maneira análoga a descrita aqui,

sendo ausente apenas este último processo de identificação e comparação. O uso de uma SVM para a classificação é uma prática que vem crescendo em ciências biológicas, devido ao grande tamanho de suas bases de dados (SCHOELKOPF; TSUDA; VERT, 2004).

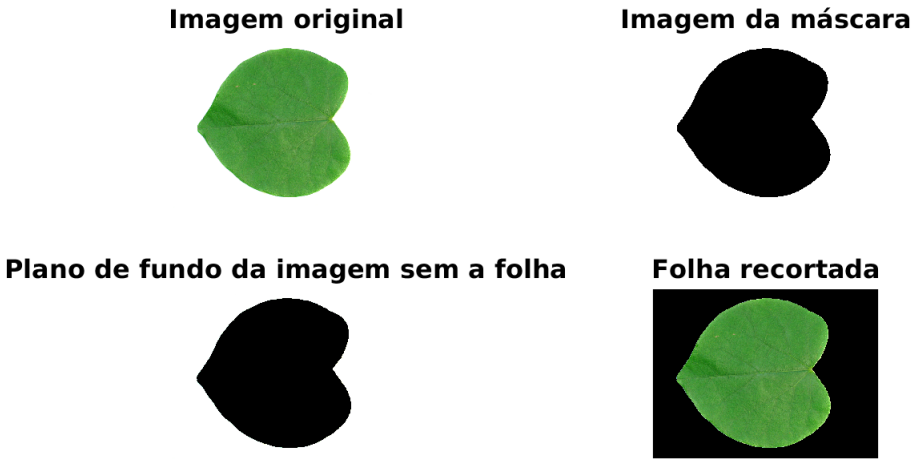
3.3 Pré-processamento

Por questões de conveniência e unificação de padrões, as imagens da base de dados, com um plano de fundo branco, são processadas para a remoção de algum ruído ou distorção e troca para um plano de fundo preto. Isso é feito através da análise dos canais de cor das imagens.

As folhas da base de dados variam apenas entre os tons verde, amarelo e marrom. Em relação às cores das folhas, o fundo branco utilizado na base de dados apresenta valores no canal azul muito mais altos que os valores das folhas em si. Assim, a seleção da folha pode ser feita a partir da imposição de um limiar no canal azul. Testes para diferentes valores de limiar foram feitos, e chegou-se a conclusão que um limiar de 150 no valor do canal azul permite a eliminação do plano de fundo branco.

Após essa etapa, algumas pequenas imperfeições na imagem (sujeira, alguns pedaços de folhas) presentes na base de dados ainda podem atrapalhar a análise. Assim, é feita uma detecção do maior corpo presente na imagem: a folha. Apenas esse corpo será preservado na imagem final. Para fins de padronização, a folha assim recortada é superposta sobre um plano de fundo preto. A Figura 17 exibe a conversão desse processo.

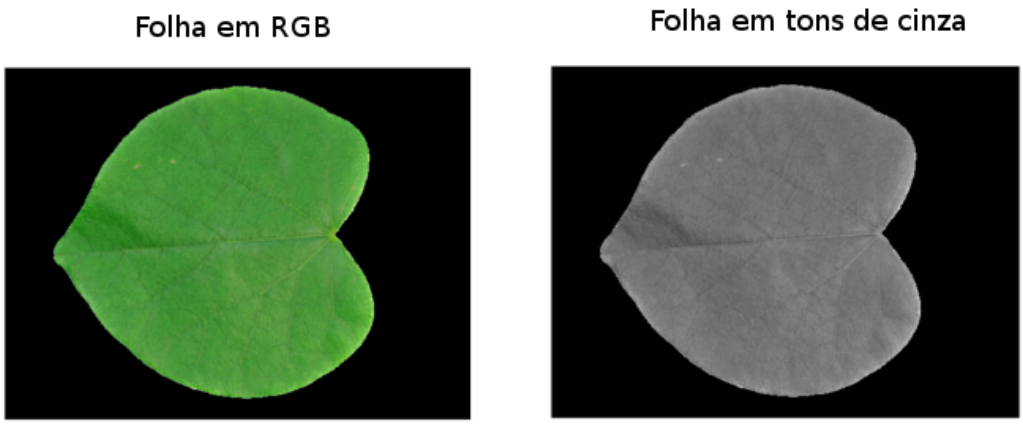
Figura 17: Uma folha da base de dados com o plano de fundo preto.



Fonte: autoria própria.

O algoritmo SIFT requer como entrada uma imagem em tons de cinza. Utilizando a Equação 2.1, imagem em 3 canais RGB é convertida para uma imagem em tons de cinza. O resultado encontra-se na Figura 18.

Figura 18: Exemplo de conversão de uma imagem RGB para uma imagem em tons de cinza.



Fonte: autoria própria.

3.4 Aplicação do algoritmo SIFT

O algoritmo SIFT como descrito em (LOWE, 2004) seleciona automaticamente as regiões de interesse que possuirão descritores. Um exemplo de tal seleção está na Figura 19.

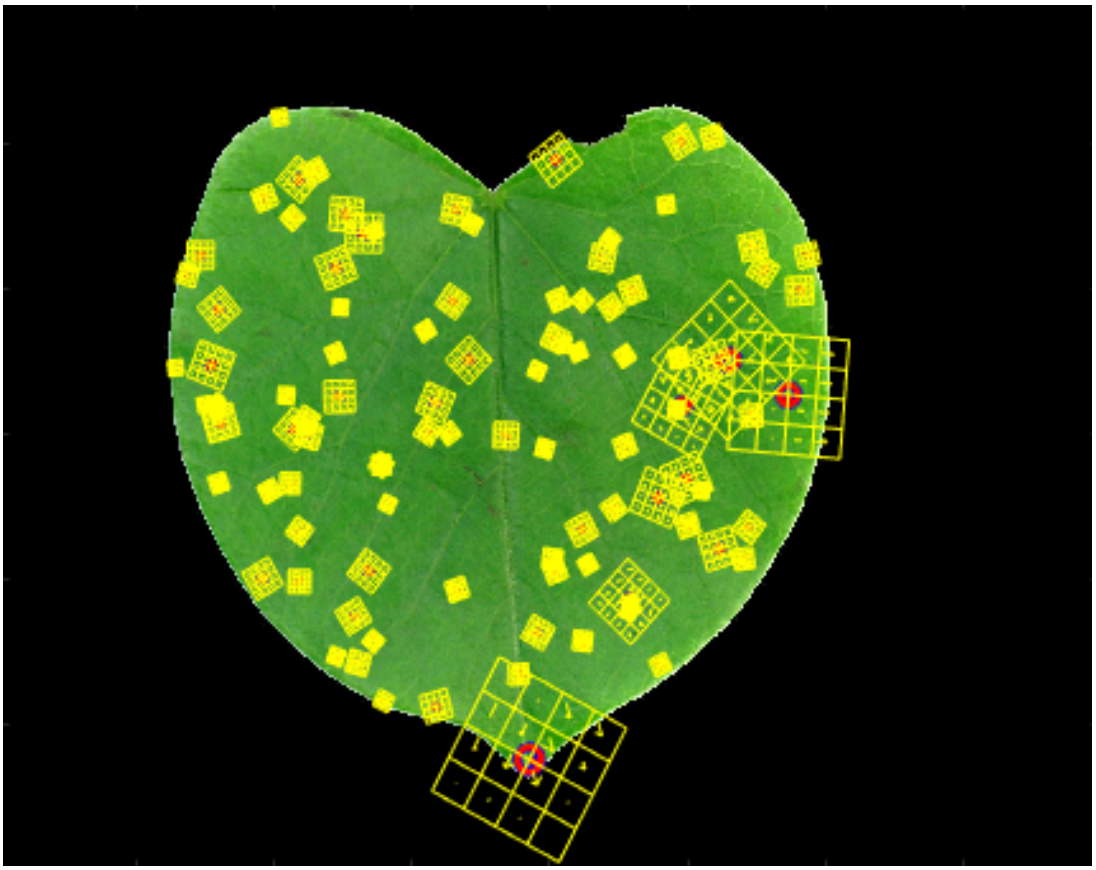
Uma consequência desse fato é que a análise, por esse algoritmo, de cada folha resultará em um número diferente de descritores. Algumas espécies possuem mais descritores e outras menos; além disso, as regiões da folha que esses descritores foram retirados também variam em cada imagen.

A fim de realizar uma comparação entre as imagens da base de treinamento e da base de testes, optou-se pela aplicação do SIFT não seguindo às regras o algoritmo de Lowe, mas sim na criação de uma malha nas imagens. Assim, a criação dos descritores não será mais vinculada às regiões de interesse do algoritmo, e sim a regiões predeterminadas da imagem, garantindo que todas as imagens possuam o mesmo número de descritores.

(BOSCH; ZISSERMAN; MUÑOZ, 2006) e (BOSCH; ZISSERMAN; MUÑOZ, 2007) mostram que o cálculo de descritores por tal método geralmente traz uma classificação melhor, consolidando o método. É importante notar que a análise em diferentes escalas e de rotação ainda permanece; apenas o local de obtenção dos descritores que muda.

A aplicação do algoritmo SIFT foi, então, feita em uma malha com divisões de 40x40 pixels. A criação dos descritores foi feita utilizando 4 escalas diferentes: janelas de 4x4, 6x6, 8x8 e 10x0 pixels. Assim, para uma imagem de 1600x1200 pixels, como são as

Figura 19: Amostra de 100 *keypoints* obtidos na análise de uma folha.



Fonte: autoria própria.

do projeto FLAVIA, cada escala possui 1200 descritores. As 4 escalas unidas resultarão em 4800 descritores por imagem. Um exemplo da malha aplicada a imagem da folha está na Figura 20.

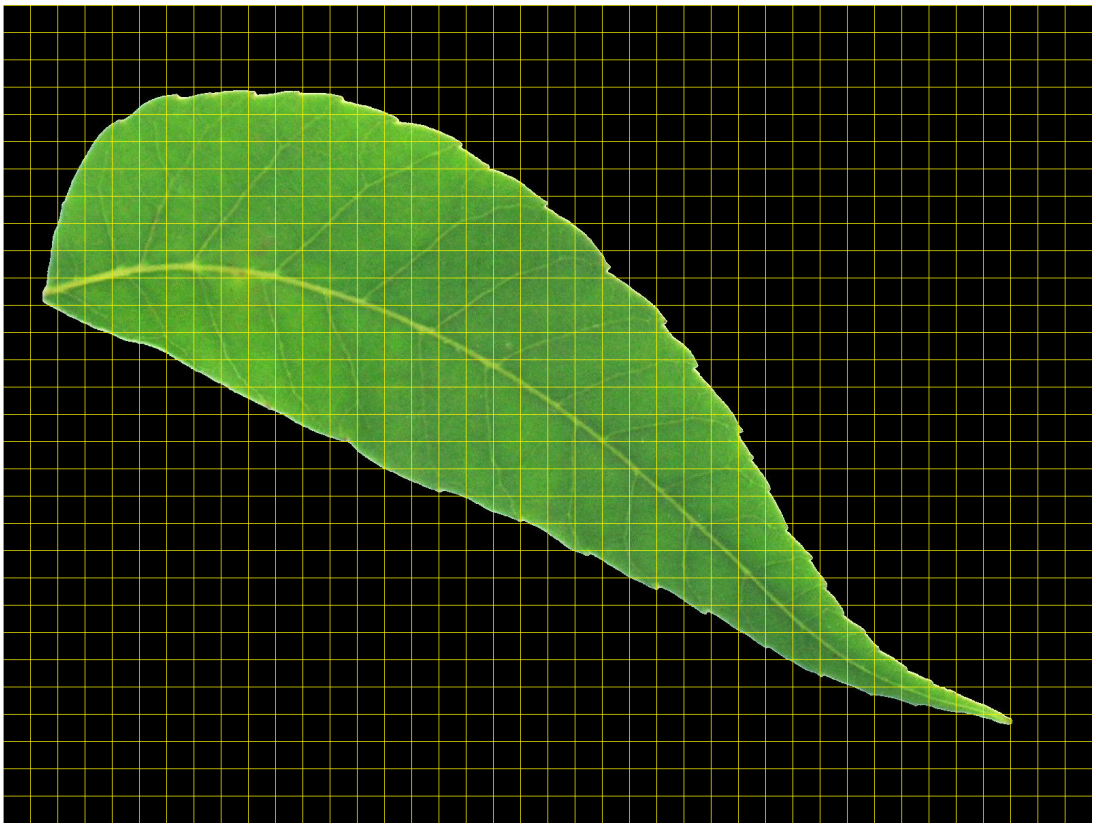
Para fins de comparação, a aplicação do SIFT sem a malha, na base de dados, resultou em um total de 2987329 descritores. Com a implementação da malha, o número quase dobrou: foram gerados 5376000 descritores.

3.5 Cáculo dos momentos de cor e concatenação com os descritores SIFT

O mesmo princípio que foi aplicado na aplicação do SIFT em malha nas imagens é utilizado para a implementação do cálculo dos momentos de cor. Para que a concatenação com os descritores SIFT fosse possível, a malha utilizada para o cálculo dos momentos de cor é a mesma utilizada anteriormente. Os momentos de cor foram calculados a partir das equações 2.15, 2.16, 2.17 e 2.18.

Considerando que os 4 momentos são calculados nos 3 canais de cor, essa etapa resulta em uma adição de 12 novas linhas de descritores no vetor. Assim, ao ser somado com os descritores SIFT, que possuem 128 linhas, o descritor final utilizado possui 140

Figura 20: Exemplo de malha aplicada para o cálculo dos descritores



Fonte: autoria própria.

linhas.

A eficácia de tal junção foi averiguada pela comparação do desempenho do descritor SIFT sozinho, sem a junção com os momentos de cor, com a versão ampliada, com a junção com os momentos de cor.

3.6 Agrupamento pelo método *k-means*

O algoritmo *k-means* foi utilizado com suas modificações mencionadas anteriormente: o algoritmo de Elkan e a inicialização *k-means++*. Para que a comparação supracitada fosse realizada, os centroides foram geradas, tanto para os descritores com os momentos de cor, como para os descritores sem os momentos de cor.

A escolha do número de centroides ótimo a ser utilizado foi feita com base na análise de uma curva de desempenho de classificação em relação ao número utilizado de centroides. O levantamento de tal curva envolveu a repartição do total de descritores em grupos de 3, 5, 7, 30, 50, 70, 100, 300, 500, 700 e 1000 centroides.

A distância utilizada no cálculo do algoritmo é a distância L2, isto é, a distância euclidiana. Ela é apresentada na Equação 3.1, na qual p e q são os pontos a serem analisados

e $d(p, q)$ é a distância L2.

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + ... + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2} \quad (3.1)$$

3.7 *Bag-of-words*

Com os centroides em mãos, a classificação dos descritores em palavras visuais, ou, em outras palavras, a alocação dos descritores para o melhor centroide, é feita.

A distribuição dos descritores é realizada através da análise de sua distância aos centroides. O descritor será alocado ao centroide que ele possuir menor distância, formando os conjuntos de palavras visuais. A distância utilizada será do mesmo tipo utilizado no cálculo dos centroides: distância L2. O resultado deste processo é um histograma que representa a frequência de ocorrência de cada palavra visual em uma determinada imagem.

3.8 Treinamento da SVM

A SVM recebe como entrada o histograma gerado no passo anterior e um vetor que mostra as classes reais que eles pertencem, ou seja, cada imagem analisada geram um histograma, que é vinculado a uma classe, representando a espécie à qual a folha analisada pertence. O treinamento da SVM é feito utilizando uma função de *kernel* linear, o próprio produto escalar.

Após a etapa de treinamento, ao ser alimentada com um outro histograma resultante da análise da imagem, a SVM deverá realizar a classificação da entrada comparando-a com os outros vetores analisados quando do treinamento.

3.9 Identificação da base de testes

Nessa etapa, o objetivo é a classificação de uma imagem pertencente à base de testes pela SVM já treinada, descrita anteriormente. Para isso, processos similares são realizados na base de testes. Todas as etapas são executadas seguindo a mesma ordem e parâmetros mencionados acima, incluindo o processamento dos descritores com e sem os momentos de cor.

A única diferença é que, ao invés de utilizar os histogramas gerados para o treinamento da SVM, eles servirão de entrada para a SVM treinada. A SVM, então, realizará sua classificação.

4 RESULTADOS E DISCUSSÃO

Levando em consideração os fatos mencionados anteriormente, os resultados aqui expostos abarcam tanto os descritores SIFT concatenados com os momentos de cor, quanto os descritores SIFT puros.

4.1 Descritores SIFT puros

4.1.1 Extração dos descritores SIFT

Conforme mencionado no capítulo anterior, uma imagem da base de dados, de tamanho 1600x1200, gera um vetor 128x4800, contendo os descritores SIFT analisado na malha. Os descritores gerados por todas 1354 imagens, para a aplicação da clusterização pelo método *k-means*, foram unidos em um único vetor com dimensões de 128x6499200.

4.1.2 *k-means*

A determinação do melhor número de centroides é feita pelo método do cotovelo, cuja ideia central foi primeiramente exposta em (THORNDIKE, 1953). O método consiste em escolher um número de centroides a partir do qual não haja mais variação significante no resultado de classificação dos dados.

A Figura 21 mostra a eficácia da classificação para ensaios com 3, 5, 7, 10, 30, 50, 70, 100, 300, 500, 700 e 1000 centroides, em escala linear. A Figura 22 exibe os mesmos dados, mas em escala logarítmica, permitindo mua melhor visão na região inicial, onde ocorre maior variação de eficácia. A tabela 2 mostra os valores encontrados em cada ponto analisado.

É possível notar que a variância na taxa de acertos possui sua maior variação, em função do número de centroides, no intervalo de 3 a 30 centroides escolhidos – nesse intervalo, a taxa varia de 24,37% até 66,67%, isto é, uma variação de 43,2% na taxa de acertos. No intervalo de 50 até 1000 centroides, a taxa varia muito pouco, atingindo um pico de 77,71% em 500 centroides – exibindo uma variação de apenas 5,83%.

Assim, dentre os pontos levantados, a melhor taxa de acertos oferecida é quando 500 centroides são utilizados. Entretanto, o aumento da taxa é substancialmente menor para cada centroide adicionado a partir de 30 centroides.

Nota-se, ainda, que a relação entre a taxa de acertos e o número de centroides não é uma relação monótona crescente. Embora a tendência da taxa seja crescer com o aumento do número de centroides, isso nem sempre ocorre. Os pontos com 70, 700, e 1000 centroides mostram que pode existir uma leve queda na taxa de acertos entre números próximos de centroides utilizados.

Tabela 2: Acurácia em relação ao número de centroides utilizados.

Acurácia (%)	Número de centroides
24,37	3
41,04	5
43,54	7
63,12	10
66,67	30
71,88	50
71,67	70
73,96	100
77,08	300
77,71	500
74,38	700
74,17	1000

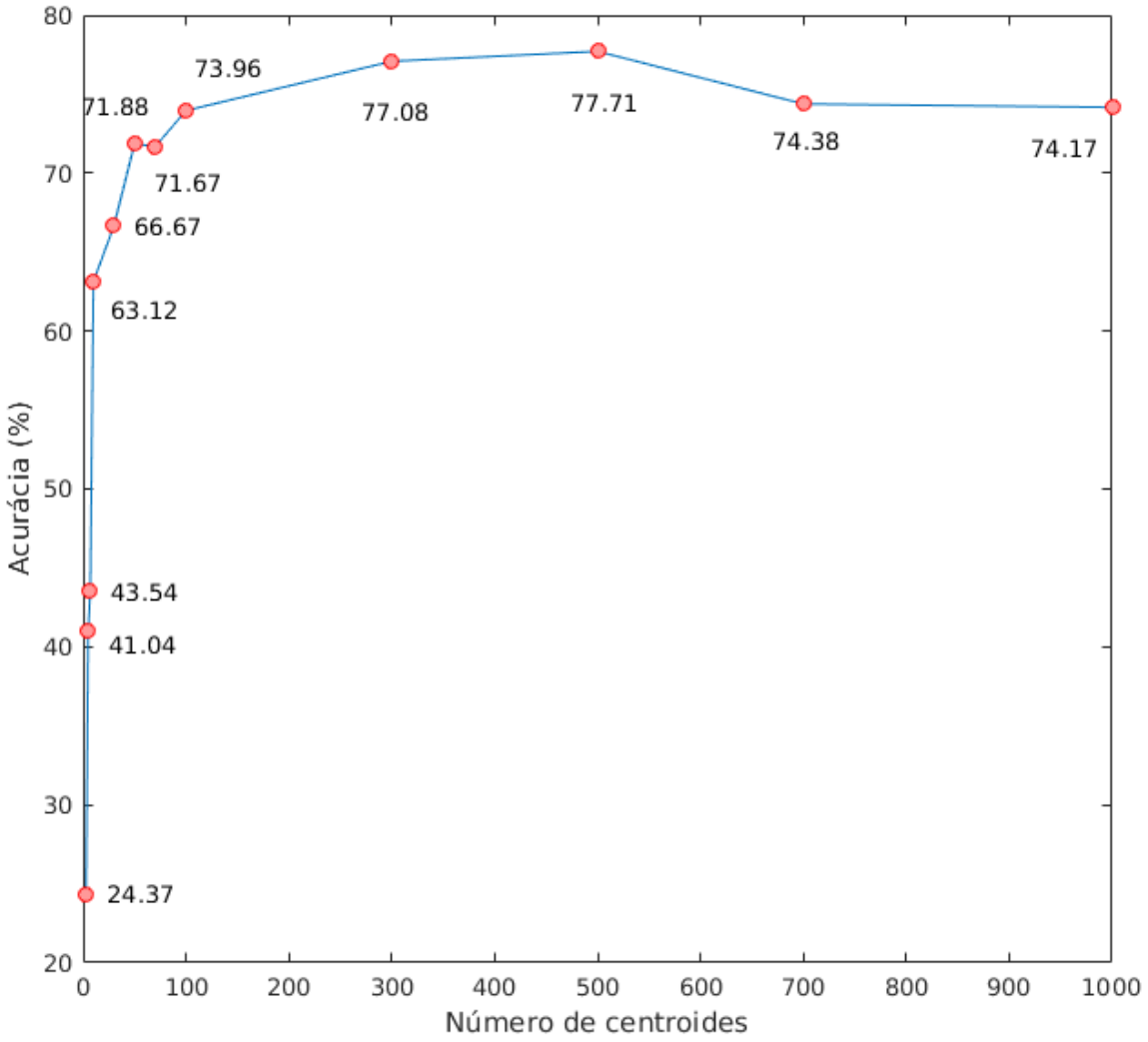
4.1.3 Histogramas e Bag-of-words

A classificação de cada descritor computado em algum cluster definido no passo anterior gera um histograma de frequência de palavras visuais para cada descritor. O conjunto desses histogramas com os dados da base de treinamento servirá como entrada para o treinamento da SVM. A Figura 24 mostra tal histograma, para a folha de espécie Lagerstroemia indica (L.) Pers., retirada do projeto FLAVIA, com nome de arquivo igual a "2520.jpg". A folha é exibida na Figura 23. O número de centroides escolhidos foi de 500, para obter a maior acurácia dentre os casos testados: é esse também o número das palavras visuais.

No histograma da Figura 24, percebe-se a o primeiro cluster possui, em comparação com os outros, uma quantidade maior de ocorrências. Isso se deve ao fato da inserção da análise do SIFT em malhas: nesse caso, não ocorre a descrição somente de pontos de importância elevada, mas se de qualquer célula da malha. As imagens do projeto FLAVIA não são compostas somente das folhas: há um plano de fundo branco. No pré-processamento esse plano de fundo foi filtrado e transformado para preto, mas ainda existe um plano de fundo. A alta frequência do cluster de número 1 reflete a presença do plano de fundo preto na imagem.

Como todas as imagens possuem essa mesma característica, a *variação* nesse cluster, possuindo ele maior número de ocorrências ou não, é o que influenciará a SVM na identificação de classes, assim como ocorre para cada outro cluster nos histogramas dos descritores.

Figura 21: Gráfico em escala linear da eficácia de classificação em função do número de centroides utilizados.



Fonte: autoria própria.

4.1.4 SVM

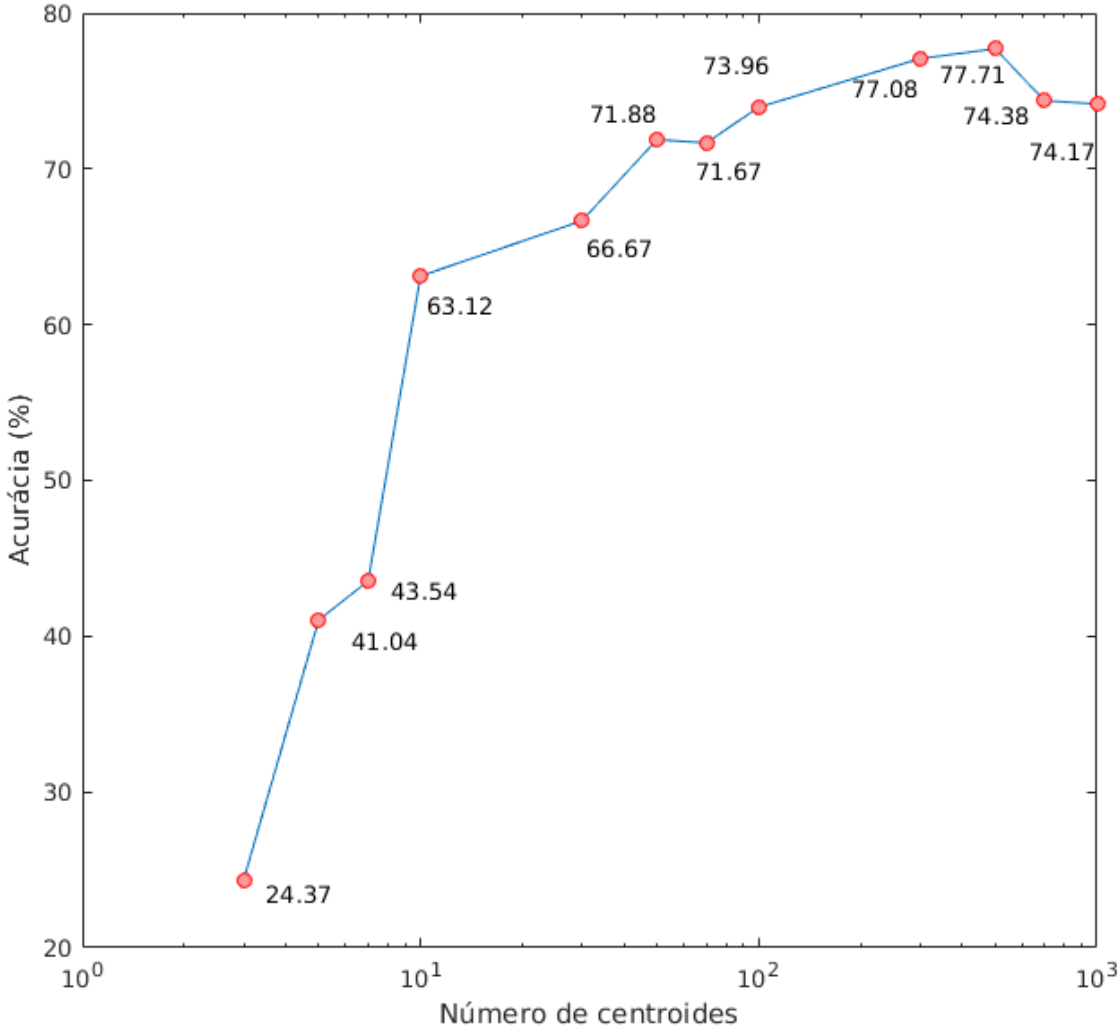
A SVM recebe como entrada o conjunto dos histogramas descritos na seção anterior e um arquivo de classificação, isto é, a qual espécie cada histograma pertence. Após isso, a SVM tenta classificar cada foto da base de testes.

As fotos da base de teste são processadas da mesma maneira que as fotos da base de dados: cada uma das 480 fotos possui um vetor de descritores de dimensões de 128x4800.

Para o caso de maior taxa de acertos, quando o histograma possue 500 palavras visuais, a SVM gerou a matriz de confusão das Figura 25. A tabela **??** mostra a relação entre o número da classe e o nome da espécie correspondente.

A Figura 25 exibe a matriz de confusão gerada pela análise da base de testes

Figura 22: Gráfico em escala logarítmica da eficácia de classificação em função do número de centroides utilizados.



Fonte: autoria própria.

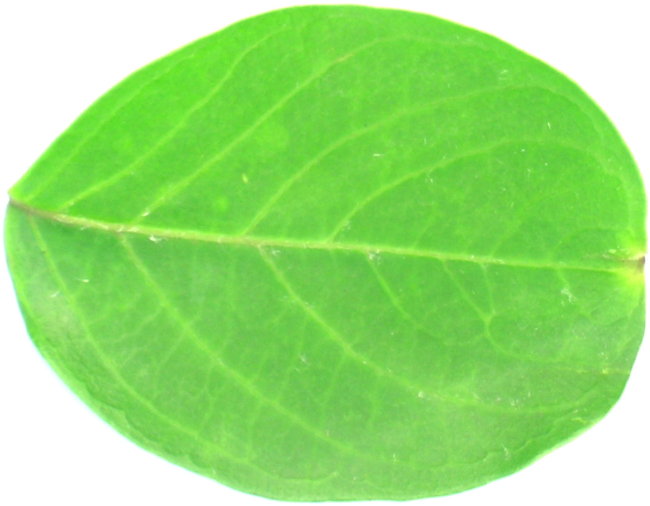
por meio do algoritmo SIFT puro. Nessa matriz, cada coluna representa a classificação verdadeira de cada imagem analisada. Cada classe foi validada com o uso de 15 imagens. As linhas representam a previsão das classes feitas pela SVM. A diagonal mostra quantos acertos a SVM teve em cada classe. Tal dado é explicitado em forma percentual na coluna à direita.

Analisando a matriz de confusão da Figura 25 e levantando dados sobre os verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, pode-se levantar algumas observações:

Em relação aos verdadeiros negativos:

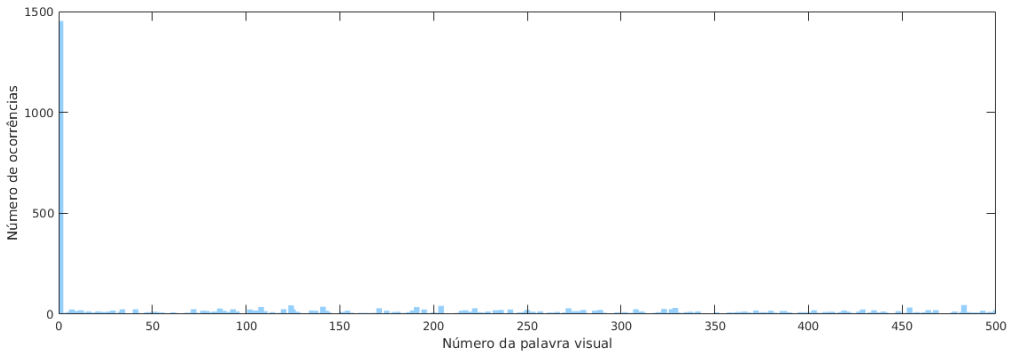
- As classes 3, 4, 8, 9, 15, 17, 18, 20, 23, 28, e 31 (total de 11 classes) possuem 100%

Figura 23: Folha de Lagerstroemia indica (L.) Pers., usada para o cálculo do histograma da Fig. 24



Fonte: adaptada de (WU et al., 2007).

Figura 24: Histograma de palavras visuais de uma folha de Lagerstroemia indica (L.) Pers., retirada de (WU et al., 2007).



Fonte: autoria própria.

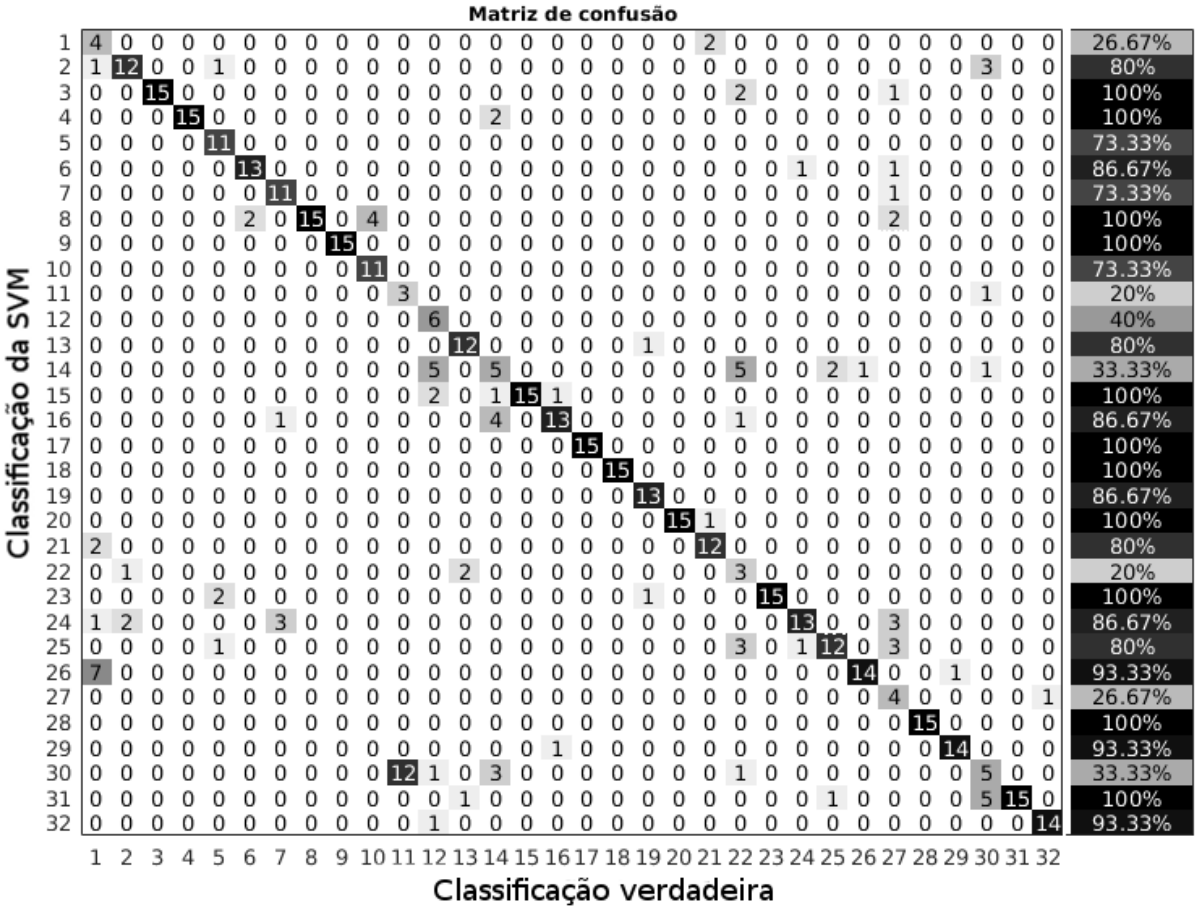
de taxa de verdadeiros negativos;

- a classe 2 apresentou a menor porcentagem para os verdadeiros negativos, com 97,47%;

- a taxa de identificação como verdadeiro negativo obteve uma média de 99,27%, com desvio padrão de 0,88%;

Em relação aos verdadeiros positivos:

Figura 25: Matriz de confusão para caso de descritores SIFT puros, com 500 palavras visuais.



Fonte: autoria própria.

- As classes 5, 9, 10, 12, 17, 18, 19, e 28 (total de 8 classes) possuem 100% de taxa de verdadeiros positivos;
- as classes 14, 22 e 30 apresentam as menores porcentagens para os positivos verdadeiros, com 26,32%, 50,00% e 22,73%, respectivamente;
- a taxa de identificação como verdadeiro positivo obteve uma média de 79,58%, com desvio padrão de 20,59%;

Em relação aos falsos positivos:

- As classes 5, 9, 10, 12, 17, 18, 19, e 28 (total de 8 classes) possuem 0% de taxa de falsos positivos;
- as classes 14, 22 e 30 apresentam as maiores porcentagens para os falsos positivos, com 73,69%, 50,00% e 77,27%, respectivamente;

- a taxa de identificação como falso positivo obteve uma média de 20,42%, com desvio padrão de 20,59%;

Em relação aos falsos negativos:

- As classes 3, 4, 8, 9, 15, 17, 18, 20, 23, 28, e 31 (total de 11 classes) possuem 0% de taxa de falsos negativos;
- a classe 22 apresenta a maior porcentagem para os falsos negativos, com 2,5%;
- a taxa de identificação como falso negativo obteve uma média de 0,73%, com desvio padrão de 0,88%;

Através desses dados, pode-se perceber que a precisão da classificação, embora siga uma guia comum, varia entre as classes. As classes 1, 11, 13, 14, 22, 27 e 30 são as que apresentam maior problema, cada uma com menos de 6 imagens identificadas corretamente. As classes 11 (Ilex macrocarpa Oliv.) e 22 (Prunus serrulata Lindl. var. lannesiana auct.) demonstraram maior dificuldade. Nota-se que a classificação entre a classe 11 (Ilex macrocarpa Oliv.) e a classe 30 Populus ×canadensis Moench) foi a que apresentou maior número de erros: 12 imagens da classe 11 foram erroneamente classificadas como pertencentes à classe 30. A figura 26 exibe as folhas das duas espécies.

As médias e seus respectivos desvios mencionados acima estão sumarizados na tabela 3.

Tabela 3: Média e desvio padrão de algumas porcentagens da matriz de confusão.

	Média (%)	Desvio padrão (%)
Verdadeiro negativo	99,27	0,88
Verdadeiro positivo	79,58	20,59
Falso positivo	20,42	20,59
Falso negativo	0,73	0,88

4.2 Descritores SIFT com momentos de cor

A computação dos descritores SIFT com os momentos de cor seguiu procedimento análogo ao mencionado acima. Entretanto, após a computação dos descritores SIFT, ocorre o concatenamento com os momentos de cor.

Em cada imagem, é computado um vetor 128x4800, composto pelos descritores SIFT. São 4 os momentos de cor, aplicados aos três canais: são, no total, 12 momentos de cor. Assim, o vetor final de descritores, em cada imagem, possui as dimensões 140x4800.

Figura 26: Imagem das duas folhas: à esquerda, folha de Ilex macrocarpa Oliv. À direita, folha de Populus x canadensis Moench.

Fonte: adaptada de (WU et al., 2007).

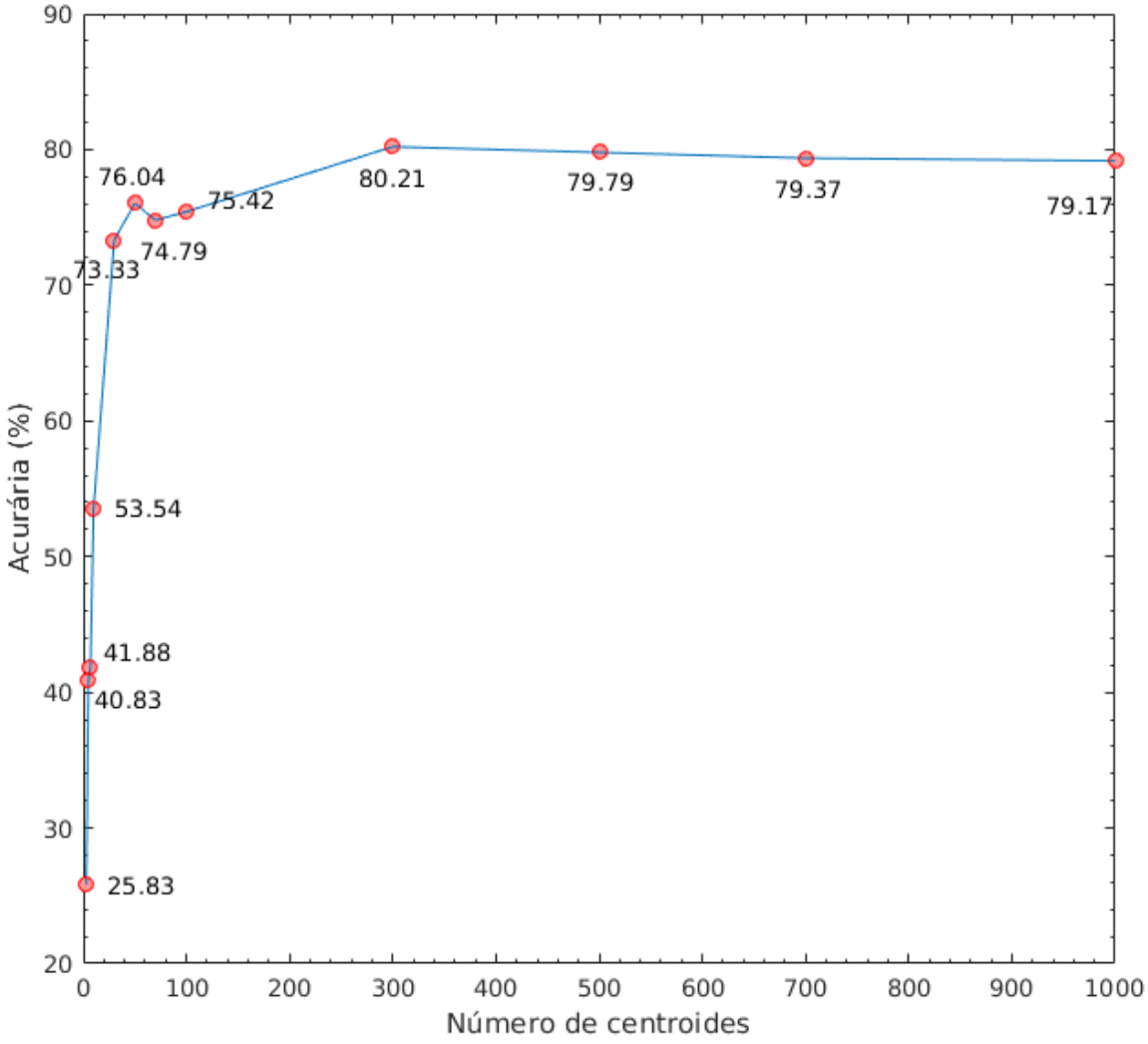
Tabela 4: Acurácia em relação ao número de centroides utilizados.

Acurácia (%)	Número de centroides
25,83	3
40,83	5
41,88	7
53,54	10
73,33	30
76,04	50
74,79	70
75,42	100
80,21	300
79,79	500
79,37	700
79,17	1000

4.2.1 *k-means*

A mesma análise para a fixação do número ótimo de centroides feita para os descritores SIFT puros é repetida aqui. As figuras 27 e 28 exibem os resultados obtidos. A tabela 4 evidencia os valores dos pontos experimentais.

Figura 27: Gráfico em escala linear da eficácia de classificação em função do número de centroides utilizados, com momentos de cor.



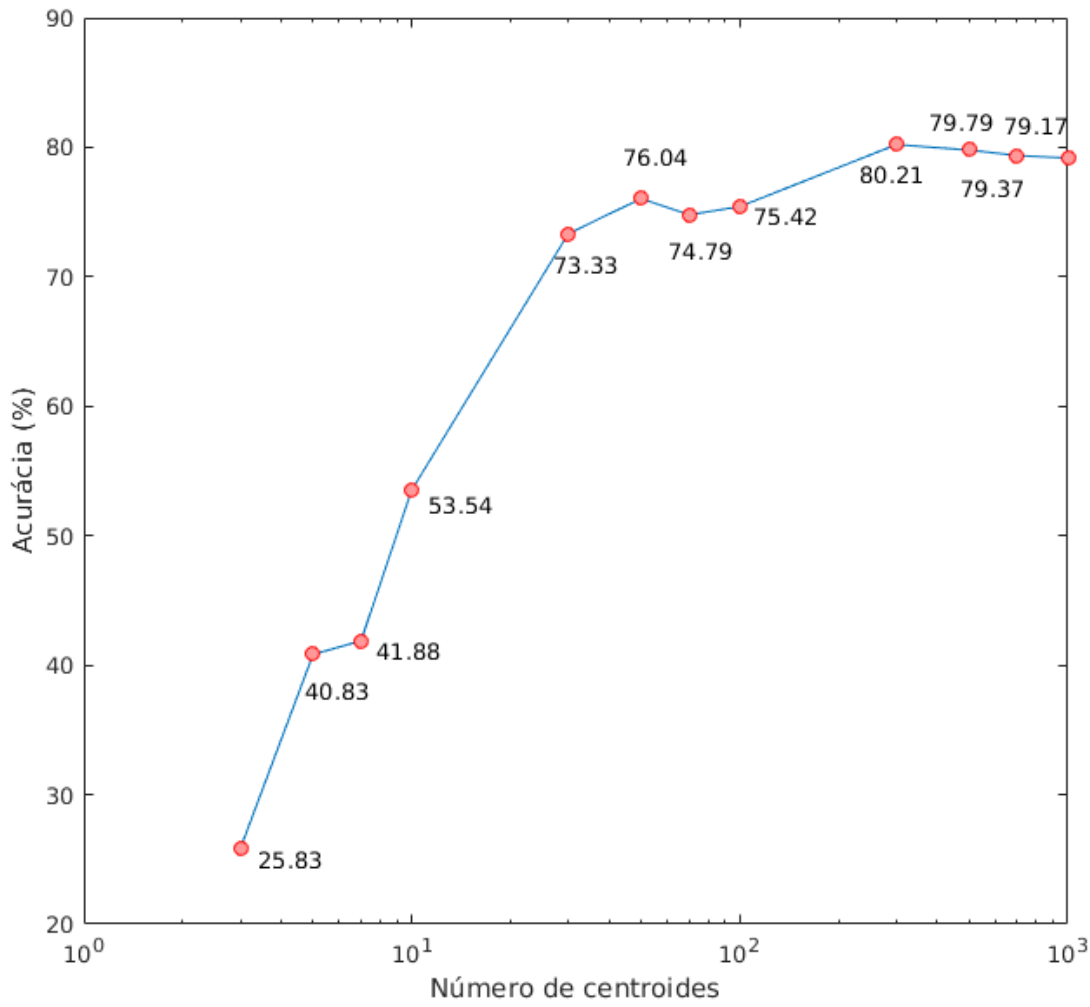
Fonte: autoria própria.

A análise das figuras 27 e 28 incita conclusões análogas ao das figuras 21 e 22, com os descritores SIFT puros.

A partir de 30 centroides, não há tanto acréscimo ao desempenho da SVM quando se adiciona mais centroides. Ainda, o crescimento não ocorre sempre com o aumento: embora siga um padrão de melhora de eficácia, alguns pontos podem possuir eficácia menores que seus vizinhos anteriores.

Os valores integrais, contudo, permitem uma comparação dos dois métodos. O método com os momentos de cor apresentou uma melhora de desempenho para 9 dos 12 números de clusters considerados. No pior caso, para 10 centroides, a adição dos momentos de cor piorou a classificação em 9,58%. Nota-se que, para 10 clusters, a taxa de acertos utilizando apenas descritores SIFT sofreu crescimento atípico: passou de 43,54%, em 7

Figura 28: Gráfico em escala logarítmica da eficácia de classificação em função do número de centroides utilizados, com momentos de cor.



Fonte: autoria própria.

clusters, para 63,12%.

Essa diferença pode ser explicada pela natureza um pouco aleatória da escolha dos centroides pelo método *k-means*, em que pode ocorrer variação. Nesse caso, pode-se teorizar que a escolha dos centroides para 5 clusters foi mais efetiva que a média, resultando em uma eficácia maior desse ponto. Esse mesmo fato explica as pequenas variações de eficácia, resultando em eficácias menores mesmo com maiores números de clusters. A *tendência* é que a eficácia melhore com a adição do número de clusters, salvo essa pequena variação.

Nas duas outras instâncias em que a adição dos momentos de cor piorou a eficácia do programa, tal efeito não foi tão significativo: para 5 e 7 clusters, a diferença foi apenas de 0,21% e 1,66%. Nota-se que todos esses pioramentos ocorreram para um número pequeno de centroides. Para valores maiores, a adição de momentos de cor melhora a classificação. Isto é, em todos os outros pontos, a adição apresentou melhora na eficácia, variando de

Tabela 5: Comparação das taxas de acertos com descritores SIFT puros e descritores com momentos de cor.

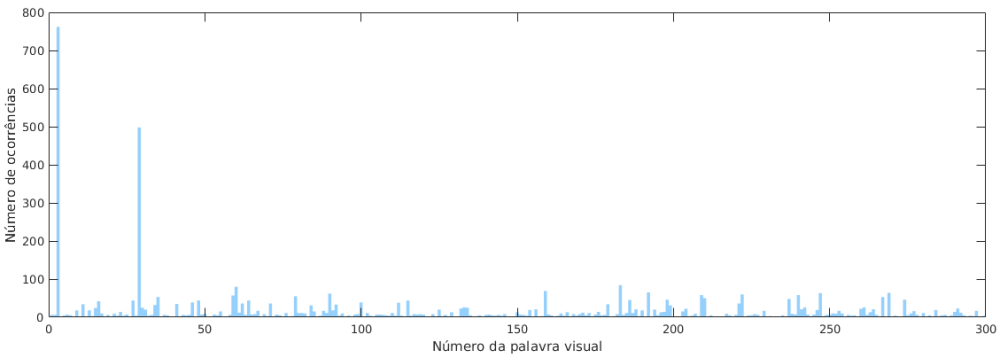
Nº de centroides	Acertos sem momentos (%)	Acertos com momentos (%)	Diferença (%)
3	24,37	25,83	1,46
5	41,04	40,83	-0,21
7	43,54	41,88	-1,66
10	63,12	53,54	-9,58
30	66,67	73,33	6,66
50	71,88	76,04	4,16
70	71,67	74,79	3,12
100	73,96	75,42	1,46
300	77,08	80,21	3,13
500	77,71	79,79	2,08
700	74,38	79,37	4,99
1000	74,17	79,17	5

1,46% até 5%. No total, a adição dos momentos resultou em uma melhora média de 1,75%, com desvio padrão de 4,25%. A tabela 5 expõe esses resultados.

4.2.2 Histogramas e Bag-of-words

A mesma folha da Figura 23 é analisada, agora com os momentos de cor. O histograma gerado pela nova classificação, agora com 300 centroides, para o melhor caso, está exposto na Figura 29.

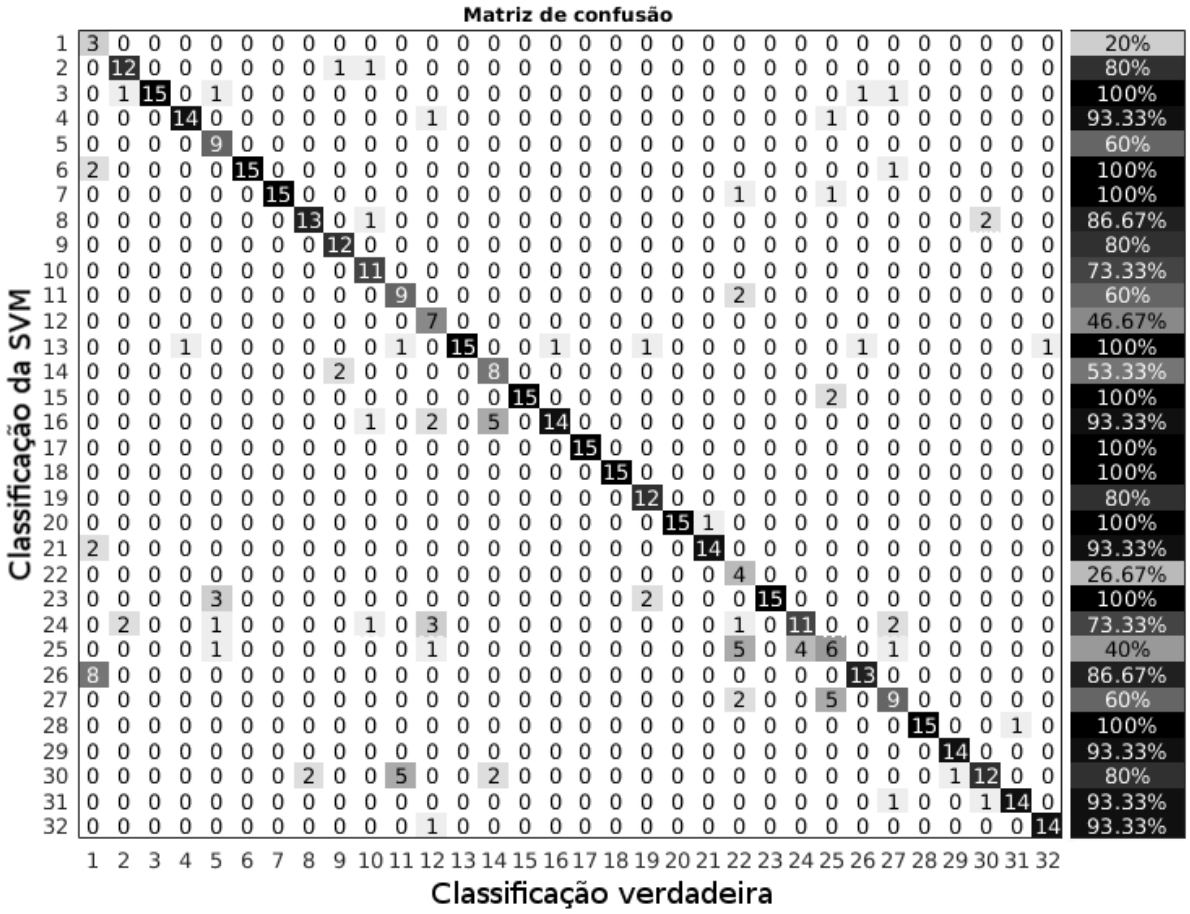
Figura 29: Histograma de palavras visuais de uma folha de Lagerstroemia indica (L.) Pers., retirada de (WU et al., 2007), com momentos de cor.



Fonte: autoria própria.

Nota-se que a predominância do primeiro cluster continua, mas tanto sua frequência, quanto a frequência dos outros clusters, mudou. Em nota, tal primeiro cluster não é o

Figura 30: Matriz de confusão para caso de descritores SIFT com momentos de cor, com 300 palavras visuais.



Fonte: autoria própria.

mesmo da análise anterior. Isso se deve à adição dos momentos, que acaba gerando uma outra divisão em centroides.

4.2.3 SVM

Novamente, a matriz de confusão, agora na Figura 30, e os respectivos dados sobre os verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos são analisadas:

Em relação aos verdadeiros negativos:

- As classes 4, 6, 7, 13, 15, 17, 18, 20, 23 e 28 (total de 10 classes) possuem 100% de taxa de verdadeiros negativos;
- a classe 1 apresentou a menor porcentagem para os verdadeiros negativos, com 97,48%;

- a taxa de identificação como verdadeiro negativo obteve uma média de 99,37%, com desvio padrão de 0,73%;

Aqui pode-se notar que o número de classes com taxa de 100% de acertos mais diminui em uma unidade: antes eram 11 tais classes, agoram são apenas 10. Entretanto, a média subiu um pouco, passando de 99,27% para 99,37%, enquanto o desvio padrão diminuiu de 0,88% para 0,73%.

Em relação aos verdadeiros positivos:

- As classes 3, 5, 9, 10, 12, 17, 18, 19, 22, e 29 (total de 10 classes) possuem 100% de taxa de verdadeiros positivos;
- as classes 24, 25 e 30 apresentam as menores porcentagens para os positivos verdadeiros, com 52,37%, 33,33% e 54,55%, respectivamente;
- a taxa de identificação como verdadeiro positivo obteve uma média de 83,73%, com desvio padrão de 17,33%;

O número de classes com taxa de 100% sofreu um acréscimo: de 8, anteriormente, passou para 10. A média subiu de 79,58% para 83,73%, e o desvio padrão diminuiu de 20,59% para 17,33%.

Em relação aos falsos positivos:

- As classes 1, 5, 9, 10, 12, 17, 18, 19, 22, e 29 (total de 10 classes) possuem 0% de taxa de falsos positivos;
- as classes 24, 25 e 30 apresentam as maiores porcentagens para os falsos positivos, com 47,62%, 66,67% e 45,45%, respectivamente;
- a taxa de identificação como falso positivo obteve uma média de 16,27%, com desvio padrão de 17,33%;

O número de classes com 0% de taxa subiu de 8 para 10. A média caiu de 20,42% para 16,27% e o desvio padrão passou de 20,59% para 17,33%.

Em relação aos falsos negativos:

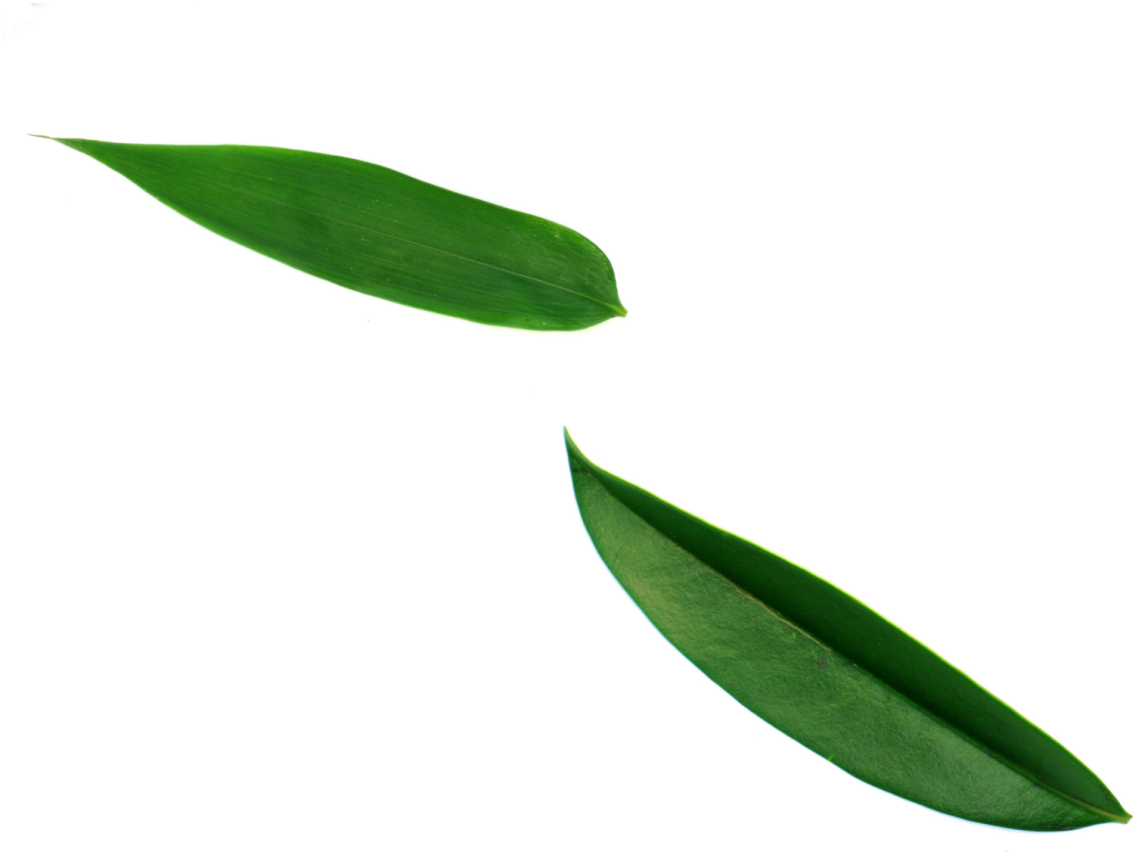
- As classes 3, 6, 7, 13, 15, 17, 18, 20, 23, e 28 (total de 10 classes) possuem 0% de taxa de falsos negativos;
- a classe 1 apresenta a maior porcentagem para os falsos negativos, com 2,52%;

- a taxa de identificação como falso negativo obteve uma média de 0,63%, com desvio padrão de 0,73%;

Neste caso, também, o número de classes com 0% de taxa caiu de 11 para 10. A média caiu de 0,73% para 0,63% e o desvio padrão passou de 0,88% para 0,73%.

Através desses dados, pode-se perceber que a precisão da classificação, embora siga uma guia comum, novamente, varia entre as classes. As classes 1, 22 e 25 são as que apresentam maior problema, com cada classe possuindo menos que 6 acertos. A classe 1 (Phyllostachys edulis (Carr.) Houz.) demonstra maior dificuldade, com apenas 3 acertos. Nota-se que a classificação entre a classe 1 (Phyllostachys edulis (Carr.) e a classe 26 (Manglietia fordiana Oliv.) foi a que apresentou maior número de erros: 8 imagens da classe 1 foram erroneamente classificadas como pertencentes à classe 26. As médias e seus respectivos desvios mencionados acima estão sumarizados na tabela 6. A figura 31 mostra as folhas das plantas das classes 1 e 26.

Figura 31: Imagem das duas folhas: à esquerda, folha de Phyllostachys edulis (Carr.) Houz. À direita, folha de Manglietia fordiana Oliv.



Fonte: adaptada de (WU et al., 2007)

Além disso, a adição dos momentos de cor gerou um crescimento nas taxas de classificação de verdadeiros positivos e negativos e uma diminuição nas taxas de falsos positivos e negativos. Em outras palavras, embora a adição dos momentos de cor não tenha melhorado de forma significativa nas classes que já apresentavam altos índices de acerto, sua adição permitiu uma melhora (de 4,15%) nas classes onde o número de acertos era mais baixo. Esses resultados estão expostos na tabela 7.

Tabela 6: Média e desvio padrão de algumas porcentagens da matriz de confusão.

	Média (%)	Desvio padrão (%)
Verdadeiro negativo	99,37	0,73
Verdadeiro positivo	83,73	17,33
Falso positivo	16,27	17,33
Falso negativo	0,63	0,73

Tabela 7: Diferenças nas porcentagens de classificação entre a análise com descritores SIFT puros e com momentos de cor.

	Diferença entre as médias (%)	Diferença entre os desvios (%)
Verdadeiro negativo	0,10	-0,15
Verdadeiro positivo	4,15	-3,26
Falso positivo	4,15	-3,26
Falso negativo	0,10	-0,15

5 CONCLUSÃO

O objetivo desse trabalho foi o desenvolvimento de método computacional para clasificar automaticamente as espécies de plantas a partir da análise da imagem da sua folha. Para isso, foram utilizados os descritores SIFT e os momentos de cor. Além da discussão sobre o tipo final de descritor a ser utilizado, também discutiu-se a quantidade ótima de palavras visuais a ser utilizada, fator que também afeta o desempenho final.

Nessas linhas gerais, a análise do método do cotovelo permitiu a definição, não de um número de clusters específicos, mas um certo número que garante que a variação entre seus vizinhos não impacta fortemente os resultados. Como visto, a partir de 30 clusters, seja na análise com ou sem momentos de cor, a eficácia não apresenta variação tão grande ao se adicionar mais clusters. A variação ainda existe, usualmente aumentando a eficácia conforme o aumento de clusters, mas, a partir desse ponto, tem seu impacto reduzido.

Já a implementação com momentos de cor mostrou-se, salvo eventuais discrepâncias já comentadas, com eficácia maior do que a implementação com apenas descritores SIFT. A maior taxa de acertos totais obtidas, no caso com os momentos de cor, com 300 palavras visuais, foi de 80,21%. O desempenho total apresentou uma melhora média de 1,75%, com desvio padrão de 4,25%. A taxa de verdadeiros positivos e negativos cresceu, enquanto a taxa de falsos positivos e negativos diminuiu. Os maiores impactos foram no crescimento da taxa de verdadeiros positivos, que passou de 79,58% para 83,73% e na diminuição da taxa de falsos positivos, que passou de 20,42% para 16,27%. Assim, a adição dos momentos de cor não influenciou tanto nas classes que já possuiam uma taxa de acertos alta, mas permitiu a melhor discriminação nas taxas que possuiam taxa de acertos baixa.

A escolha do número de descritores utilizados na análise de cada imagem possui poder de tanto limitar, quanto aumentar a eficiência. Neste trabalho, o número total de descritores utilizados, 6494400, é quase o dobro que o número de descritores calculados automaticamente (3308827) pelo algoritmo de Lowe. A utilização de ainda mais descritores pode vir a trazer resultados melhores, a custa de um maior tempo computacional. Tal custo propaga-se às diferentes etapas do programa, não ficando restrito à primeira etapa de análise de descritores.

O mesmo raciocínio poderia ser aplicado à quantidade de palavras visuais escolhidas. Novamente, o que deve se ter em mente é que, a partir de um determinado número, os ganhos obtidos não apresentarão tanta variação dos seus ganhos vizinhos. Nesses casos, apesar dos ganhos continuarem a crescer, deve-se pesar a adição de tal ganho com o crescimento do custo computacional do algoritmo.

Outros algoritmos também poderiam ser utilizados para a obtenção de descritores.

Algoritmos como o *Inner-Distance Shape Context* (IDSC), ou o *Speeded Up Robust Features* (SURF) poderiam ser utilizados no lugar do SIFT, ou em conjunto com ele. Entretanto, é necessário o que cada algoritmo analisa. O SIFT analisa características locais, que não variam com a escala, em uma imagem em tons de cinza. Além disso, ele não requer maior interação do usuário. Assim como a adição dos momentos de cor permitiu uma análise mais completa da imagem da folha da planta, justamente por trazer uma nova dimensão (a análise de cor), a adição de outros descritores pode melhorar a acurácia do programa. O uso de descritores de fronteira, em conjunto com os métodos desenvolvidos, permitiria uma maior criteriosidade na decisão, principalemte nos casos em que a SVM se confundiu. Nesses casos, embora os descritores SIFT e os momentos de cor sejam similiares, os descritores de fronteira permitiriam diferenciar uma espécie da outra.

(WÄLDCHEN; MÄDER, 2017) faz um levantamento da acurácia de outros métodos para a identificação de plantas. Dentre esses métodos, o método desenvolvido neste trabalho possui, para seu melhor caso, acurácia superior a 8 dos métodos elencados. Nota-se que os métodos que obtiveram maiores acurácias foram justamente os que realizaram a análise da imagem da folha por meio de várias maneiras. Utilizaram descritores de fronteira, de região, de textura e de cor, cada um complementando o outro. Além disso, para o caso de previsão, outros métodos além da SVM foram utilizados, mas a Máquina de Vetores de Suporte foi o método mais popular, embora fossem utilizadas diferentes funções de *kernel*.

Por fim, levando em consideração as discussões anteriores, a adição da análise por momentos de cor levou a um melhoramento no método desenvolvido. Sucessivas melhorias poderiam ser atingidas através da adição de novos descritores em conjunto com os aqui apresentados. Um descritor de borda permitiria sanar a indecisão da SVM para determinados casos onde tanto o SIFT quanto os momentos de cor não conseguem realizar a identificação. Um descritor de textura também adicionaria uma outra dimensão à análise.

REFERÊNCIAS

ARTHUR, D.; VASSILVITSKII, S. k-means++: The advantages of careful seeding. In: SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS. **Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms**. [S.l.], 2007. p. 1027–1035.

BELHUMEUR, P. et al. Searching the world's herbaria: A system for visual identification of plant species. **Lecture Notes in Computer Science**, Springer-Verlag Berlin, v. 5305, p. 116–129, 2008.

BONNET, P. et al. Plant identification: Man vs. machine. **Multimedia Tools and Applications**, Springer Verlag, v. 75 (3), p. 1647–1665, 2016.

BOSCH, A.; ZISSERMAN, A.; MUÑOZ, X. Scene classification via plsa. **Computer Vision–ECCV 2006**, Springer, p. 517–530, 2006.

_____. Image classification using random forests and ferns. In: IEEE. **Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on**. [S.l.], 2007. p. 1–8.

BOSWELL, D. Introduction to support vector machines. 2002.

C. V. Starr Virtual Herbarium. **C. V. Starr Virtual Herbarium**. 2017. Disponível em: <http://sweetgum.nybg.org/science/vh/>.

DAS, A. et al. Clearedleavesdb: an online database of cleared plant leaf images. 2014.

ELKAN, C. Using the triangle inequality to accelerate k-means. In: **Proceedings of the 20th International Conference on Machine Learning (ICML-03)**. [S.l.: s.n.], 2003. p. 147–153.

FEI-FEI, L.; FERGUS, R.; TORRALBA, A. **Recognizing and Learning Object Categories**. 2009. Disponível em: <http://people.csail.mit.edu/torralba/shortCourseRLOC/>.

GASTON, K.; O'NEILL, M. Automated species identification: why not? **Phil. Trans. R. Soc. Lond.**, London, v. 359, p. 655–667, 2004.

GASTON, K. J.; MAY, R. M. Taxonomy of taxonomists. **Nature**, v. 356, p. 281–281, 1992.

GONZALES, G. L. G. **APLICAÇÃO DA TÉCNICA SIFT PARA DETERMINAÇÃO DE CAMPOS DE DEFORMAÇÕES DE MATERIAIS USANDO VISÃO COMPUTACIONAL**. 2011. Tese (Doutorado), 2011. Disponível em: <https://doi.org/10.17771/pucrio.acad.17050>.

JAMIL, N. et al. Automatic plant identification: Is shape the key feature? **Procedia Computer Science**, v. 76, p. 436 – 442, 2015. ISSN 1877-0509. 2015 IEEE International Symposium on Robotics and Intelligent Sensors (IEEE IRIS2015). Disponível em: <http://www.sciencedirect.com/science/article/pii/S1877050915037886>.

JIN, T. et al. A novel method of automatic plant species identification using sparse representation of leaf tooth features. **PLOS ONE**, Public Library of Science (PLoS), v. 10, n. 10, p. e0139482, oct 2015. Disponível em: <https://doi.org/10.1371/journal.pone.0139482>.

KUMAR, N. et al. Leafsnap: A computer vision system for automatic plant species identification. In: **Computer Vision – ECCV 2012**. Springer Berlin Heidelberg, 2012. p. 502–516. Disponível em: <https://doi.org/10.1007/978-3-642-33709-3_36>.

LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.

LOWE, D. G. Object recognition from local scale-invariant features. In: IEEE. **Computer vision, 1999. The proceedings of the seventh IEEE international conference on**. [S.l.], 1999. v. 2, p. 1150–1157.

_____. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, Springer Nature, v. 60, n. 2, p. 91–110, nov 2004. Disponível em: <https://doi.org/10.1023/b:visi.0000029664.99615.94>.

MANGOLD, J. **Plant Identification Basics**. 2013. Agriculture and Natural Resources (Weeds), Montana State Extension.

OLULEYE, B. et al. A neuronal classification system for plant leaves using genetic image segmentation. **British Journal of Mathematics & Computer Science**, Sciencedomain International, v. 9, n. 3, p. 261–278, jan 2015.

OPENCV. **OpenCV Tutorials**. 3.3.0. ed. [S.l.], 2017. Disponível em: <https://docs.opencv.org/3.3.0/>.

PADMAVATHI, K.; THANGADURAI, K. Implementation of RGB and grayscale images in plant leaves disease detection – comparative study. **Indian Journal of Science and Technology**, Indian Society for Education and Environment, v. 9, n. 6, feb 2016. Disponível em: <https://doi.org/10.17485/ijst/2016/v9i6/77739>.

PATIL, R. et al. Grape leaf disease detection using k-means clustering algorithm. **International Research Journal of Engineering and Technology**, IRJET, v. 03, n. 04, p. 2330–2333, apr 2016.

PYPR. **PyPR Documentation**. 0.1rc3. ed. [S.l.], 2010. Disponível em: <http://pypr.sourceforge.net/>.

SABRI, N.; IBRAHIM, Z.; ROSMAN, N. N. K-means vs. fuzzy c-means for segmentation of orchid flowers. In: **2016 7th IEEE Control and System Graduate Research Colloquium (ICSGRC)**. IEEE, 2016. Disponível em: <https://doi.org/10.1109/icsgrc.2016.7813306>.

SAITOH, T.; IWATA, T.; WAKISAKA, K. Okiraku search: Leaf images based visual tree search system. In: **2015 14th IAPR International Conference on Machine Vision Applications (MVA)**. [S.l.: s.n.], 2015. p. 242–245.

SCHMIDT, L. Digitization of herbarium specimens, a collaborative project. 2007.

SCHOELKOPF, B.; TSUDA, K.; VERT, J.-P. **Kernel Methods in Computational Biology**. [S.l.]: MIT Press, 2004. 71-92 p.

SINHA, U. **SIFT: Theory and Practice**. 2017. Disponível em: <http://www.aishack.in/tutorials/sift-scale-invariant-feature-transform-introduction/>.

SIVIC, J.; ZISSERMAN, A. Efficient visual search of videos cast as text retrieval. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 31, n. 4, p. 591–606, April 2009. ISSN 0162-8828.

SONKA, M.; HLAVAC, V.; BOYLE, R. Image pre-processing. In: **Image Processing, Analysis and Machine Vision**. Springer US, 1993. p. 56–111. Disponível em: <https://doi.org/10.1007/978-1-4899-3216-7_4>.

STRICKER, M.; ORENGO, M. Similarity of color images survival data: An alternative to change-point models. **Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases III**, v. 2420, p. 381–392, 1995.

The vPlants Project. **The vPlants Project**. 2007. Disponível em: <http://symbiota4.acis.ufl.edu/seinet/vplants/portal/index.php>.

THORNDIKE, R. L. Who belongs in the family? **Psychometrika**, Springer, v. 18, n. 4, p. 267–276, 1953.

VALLIAMMAL, N.; GEETHALAKSHMI, S. N. Plant leaf segmentation using non linear k means clustering. **International Journal of Computer Science Issues**, v. 9, n. 3, p. 212–218, 2012.

WÄLDCHEN, J.; MÄDER, P. Plant species identification using computer vision techniques: A systematic literature review. **Archives of Computational Methods in Engineering**, Jan 2017. ISSN 1886-1784. Disponível em: <https://doi.org/10.1007/s11831-016-9206-z>.

WEEKS, P. J. D. et al. Taxonomy of taxonomists. **J. Appl. Entomol.**, v. 123, p. 1–8, 1999.

WU, S. G. et al. A leaf recognition algorithm for plant classification using probabilistic neural network. **CoRR**, abs/0707.4289, 2007. Disponível em: <http://arxiv.org/abs/0707.4289>.

YU, H. et al. Color texture moments for content-based image retrieval. In: **Proceedings. International Conference on Image Processing**. [S.l.: s.n.], 2002. v. 3, p. 929–932 vol.3. ISSN 1522-4880.

ZHANG, Y.; JIN, R.; ZHOU, Z.-H. Understanding bag-of-words model: a statistical framework. **International Journal of Machine Learning and Cybernetics**, Springer Nature, v. 1, n. 1-4, p. 43–52, aug 2010. Disponível em: <https://doi.org/10.1007/s13042-010-0001-0>.